

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | |
|-----------------------------------|---|--|
| In re application of: |) | Examiner: Spector, Lorraine |
| |) | |
| David BOTSTEIN, et al. |) | Art Unit: 1647 |
| |) | |
| Application Serial No. 09/991,073 |) | Confirmation No: 4049 |
| |) | |
| Filed: November 14, 2001 |) | Attorney's Docket No. 39780-2730 P1C15 |
| |) | |
| For: PRO809 POLYPEPTIDES |) | Customer No. 35489 |

EXPRESS MAIL LABEL NO. : EV 582 623 710 US
DATE MAILED: DECEMBER 22, 2005

ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES
APPELLANTS' BRIEF

MAIL STOP APPEAL BRIEF - PATENTS

Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Dear Sir:

This Appeal Brief, filed in connection with the above captioned patent application, is responsive to the Final Office Action mailed on February 23, 2005. A Notice of Appeal was filed herein on July 22, 2005. A request for a three month extension of time is requested herewith. Appellants hereby appeal to the Board of Patent Appeals and Interferences from the final rejection in this case.

The following constitutes the Appellants' Brief on Appeal.

12/28/2005 NNGUYEN1 00000079 081641 09991073

01 FC:1402 500.00 DA
02 FC:1253 1020.00 DA

I. REAL PARTY IN INTEREST

The real party in interest is Genentech, Inc., South San Francisco, California, by an assignment of the parent application, U.S. Patent Application Serial No. 09/941,992 recorded November 16, 2001, at Reel 012176 and Frame 0450.

II. RELATED APPEALS AND INTERFERENCES

The claims pending in the current application are directed to a polypeptide referred to herein as "PRO809." There exist two related patent applications, (1) U.S. Patent Application Serial No. 09/990,726, filed November 14, 2001 (containing claims directed to antibodies to the PRO809 polypeptide), and (2) U.S. Patent Application Serial No. 09/989,721, filed November 19, 2001 (containing claims directed to nucleic acids encoding PRO809 polypeptides). These two related applications are also under final rejection from the same Examiner and based upon the same outstanding rejection, therefore appeal of these final rejections are being pursued independently and concurrently herewith.

III. STATUS OF CLAIMS

Claims 122-126 and 129-131 are in this application.

Claims 1-121 and 127-128 have been canceled.

Claims 122-126 and 129-131 stand rejected and Appellants appeal the rejection of these claims.

A copy of the rejected claims in the present Appeal is provided as Appendix A.

IV. STATUS OF AMENDMENTS

All previous amendments have been entered. A copy of the rejected claims in the present Appeal is provided as Appendix A.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention claimed in the present application is related to an isolated polypeptide comprising the amino acid sequence of the polypeptide of SEQ ID NO:223, referred to in the present application as "PRO809." The PRO809 gene was shown for the first time in the present application to be significantly amplified in human lung cancers as compared to normal, non-cancerous human tissue controls (Example 170). This feature is specifically recited in

Claim 124, and carried by all claims dependent from Claim 124. In addition, the invention also claims the amino acid sequence of the polypeptide of SEQ ID NO:223, lacking its associated signal-peptide; or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025 (Claims 124-126 and 129). The invention is further directed to polypeptides having at least 95% or 99% amino acid sequence identity to the amino acid sequence of the polypeptide of SEQ ID NO:223; the amino acid sequence of the polypeptide of SEQ ID NO:223, lacking its associated signal peptide; or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025, wherein the nucleic acid encoding said polypeptide is amplified in lung tumor (Claims 122-123). The invention is further directed to a chimeric polypeptide comprising one of the above polypeptides fused to a heterologous polypeptide (Claim 130), and to a chimeric polypeptide wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin (Claim 131). PRO polypeptide variants having at least about 95-99% amino acid sequence identity with a full length PRO polypeptide sequence, or a PRO polypeptide sequence lacking the signal peptide are generally described in the specification at, for example, page 305, line 23 onwards, and percent amino acid sequence identity determination is generally described at least at, for example, pages 306-308, line 14 onwards. The preparation of chimeric PRO polypeptides (Claims 130 and 131), including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin, is set forth in the specification at page 374, lines 24 to page 375, line 9. Examples 140-143 and page 376, line 12 onwards describe the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells.

The amino acid sequence of the native "PRO809" polypeptide and the nucleic acid sequence encoding this polypeptide (referred to in the present application as "DNA57836-1338") are shown in the present specification as SEQ ID NOs:223 and 222, respectively, and in Figures 151 and 150, described on pages 295, lines 3-5. The full-length PRO809 polypeptide having the amino acid sequence of SEQ ID NO:223 is described in the specification at, for example, on page 18 and pages 140-142 and the isolation of cDNA clones encoding PRO809 of SEQ ID NO:223 is described in Example 64, page 454-455 of the specification. The specification discloses that various portions of the PRO809 polypeptide possess significant

sequence similarity to the heparan sulfate proteoglycans family of proteins and to endothelial cell adhesion molecule -1 (see, for example, page 18, page 140-142 and Example 64, page 454-455).

Finally, Example 170, in the specification at page 539, line 19, to page 555, line 5, sets forth a 'Gene Amplification assay' which shows that the PRO809 gene is amplified in the genome of certain human lung cancers (see Table 9A, page 550). The profiles of various primary lung tumors used for screening the PRO polypeptide compounds of the invention in the gene amplification assay are summarized on Table 8, page 546 of the specification.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

1. Whether Claims 122-126 and 129-131 satisfy the utility/ enablement requirement under 35 U.S.C. §§101/112, first paragraph.
2. Whether Claims 122-124 and 130-131 satisfy the written description requirement under 35 U.S.C. §112, first paragraph.
3. Whether Claims 122-126 and 129-131 are patentable under 35 U.S.C. §103(a) over Clone H74302 (EST isolated by L. Hillier *et al.*, 1995), in view of Sibson *et al.*

VII. ARGUMENTS

Summary of the Arguments

Issue 1: Utility/ Enablement

Appellants rely upon the gene amplification data of the PRO809 gene for patentable utility of the PRO809 polypeptides. This data is clearly disclosed in the instant specification in Example 170 which discloses that the gene encoding PRO809 showed significant amplification, ranging from **2.070 to 3.053** fold in different lung primary tumors. Therefore, such a gene is useful as a marker for the diagnosis of lung cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

The Examiner asserted on page 2 of the Final Office Action mailed February 23, 2005 that amplification of the PRO809 polynucleotide does not impart a specific, substantial, and credible utility to the PRO809 polypeptide since, "(a) no significant association between the claimed nucleic acids and lung cancer, such as would lend diagnostic utility, has been disclosed or established, and (b) even if such had been established, it would not be predictable that the

claimed protein would be similarly diagnostic.” The Examiner maintained the rejection based on Sen *et al.* and further cited the reference Hu *et al.* in this rejection.

Appellants submit that the Declaration of Dr. Audrey Goddard submitted with Appellants' Response filed November 4, 2004, explains that a gene identified as being amplified at least 2-fold by the disclosed gene amplification assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy. Therefore, the gene amplification levels of **2.070 to 3.053** fold for lung primary tumors is considered significant.

Appellants further submit that, even if the amplification of the PRO809 gene were due to aneuploidy (which Appellants expressly do not concede with), the art exemplified by the Hittelman *et al.* reference and Sen *et al.* as well, still supports the Appellants' position that, whether aneuploidy is a feature of cancerous, pre-cancerous tissue or damaged tissue, it still provides at least one utility for the PRO809 gene as a marker for cancer or precancerous cells or damaged tissue. Accordingly, the PRO809 gene finds utility as a diagnostic for cancer or for individuals at risk for developing lung cancer. Appellants further submit that the teachings of Hu *et al.* do not conclusively establish a *prima facie* case for lack of utility (as will be discussed in detail below).

Appellants also submit that one skilled in the art would find it more likely than not that PRO809 is useful as a diagnostic tool for detecting lung tumors. Appellants had submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is more likely than not that the encoded protein will be expressed at an elevated level. First, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* (made of record in Appellants' Response filed November 4, 2004) collectively teach that in general, gene amplification increases mRNA expression. Second, the Declaration of Dr. Paul Polakis (made of record in Appellants' Response filed November 4, 2004), principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, shows that, in general, there is a correlation between mRNA levels and polypeptide levels. Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip arrays in 2004. Clearly, the research community believes that the information obtained from these chips is useful (*i.e.*,

that it is more likely than not informative of the protein level). Therefore, one skilled in the art would find it more likely than not that PRO809 is useful as a diagnostic tool for detecting lung tumors.

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is generally a positive correlation between DNA, mRNA, and polypeptide levels, these instances are exceptions rather than the rule. In the majority of amplified genes, as exemplified by Orntoft *et al.*, Hyman *et al.*, Pollack *et al.*, the Polakis Declaration, the teachings in the art overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. The widespread, art accepted use of information obtained from array chips as diagnostic markers lends further support. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO809 gene, that the PRO809 polypeptide is concomitantly overexpressed and has utility in the diagnosis of lung cancer.

Appellants further submit that even if there were no correlation between gene amplification and increased mRNA/protein expression, (which Appellants expressly do not concede), a polypeptide encoded by a gene that is amplified in cancer would still have a specific, substantial, and credible utility. Appellants submit that, as evidenced by the Ashkenazi Declaration and the teachings of Hanna and Mornin (both made of record in Appellants' Response filed November 4, 2004), simultaneous testing of gene amplification and gene product over-expression enables more accurate tumor classification, even if the gene-product, the protein, is not over-expressed. This leads to better determination of a suitable therapy for the tumor, as demonstrated by a real-world example of the breast cancer marker HER-2/neu.

Accordingly, Appellants submit that when the proper legal standard is applied, one should reach the conclusion that the present application discloses at least one patentable utility for the claimed PRO809 polypeptides. The specification also provides detailed guidance as to how to identify and make polypeptides having at least 95% amino acid sequence identity to PRO809. Accordingly, one of ordinary skill in the art would understand how to make and use the recited polypeptide variants for the diagnosis of lung cancer without any undue experimentation.

Issue 2: Written Description

Claims 122-124 and 130-131 stand rejected under 35 U.S.C. §112, first paragraph, allegedly "as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.." (Page 7 of the Final Office Action mailed February 23, 2005).

The factors to be considered in evidencing possession of a claimed genus include "disclosure of complete or partial structure, physical and/or chemical properties, functional characteristics, structure/function correlation, methods of making the claimed product, or any combination thereof." Current applicable case law holds that biological sequences are not adequately described solely by a description of their desired functional activities. It is, however, well established that a combination of functional and structural features suffices to describe a claimed genus, as discussed in the PTO's own Written Description Guidelines, and as set forth in *Enzo Biochem., Inc. v. Genprobe, Inc.* Appellants note that the claims recite structural features, namely, 95% sequence identity to SEQ ID NO:223, which are common to the genus. The genus of claimed polypeptides is further defined by having a specific functional activity for the encoding nucleic acids, namely, that the encoding nucleic acid is amplified in lung tumors. The specification provides detailed guidance as to how to identify the recited variants of SEQ ID NO:223, including methods for determining percent identity between two amino acid sequences, as well as listings of exemplary and preferred sequence substitutions, as well as detailed protocols for determining whether a gene encoding a variant PRO809 protein is amplified in lung tumor. Thus one of skill in the art could easily identify whether a variant PRO809 sequence falls within the parameters of the claimed invention.

Accordingly, a description of the claimed genus has been achieved by the recitation of both structural and functional characteristics.

Issue 3: Anticipation by Hillier *et al.*, 1995 (EST clone H74302)

Claims 122-126 and 129-131 remain rejected under 35 U.S.C. §103(a) as being unpatentable over clone H74302 (EST isolated by Hillier *et al.*, 1995) in view of Sibson *et al.*, WO94/01548.

The instant application claims PRO809 polypeptides. The Hiller EST clone does not disclose nor reduces to practice the sequence of SEQ ID NO:223. Further, Hillier does not disclose that their EST was part of a cDNA sequence that coded for a polypeptide (see enclosed alignment Item 13, Evidence Appendix). Thus, Hillier *et al.* did not teach nor reduce to practice the instantly claimed polypeptide sequences. Further, one of skill in the art would not have been able to envision the cDNA clone defined in SEQ ID NO: 222 from the teachings of Hillier and therefore, Hillier *et al.* does not anticipate the instant invention. Nor does Sibson anticipate the instantly claimed polypeptides. Thus, the instant claims are patentable over the cited art.

These arguments are all discussed in further detail below under the appropriate headings.

Response to Rejections

ISSUE 1. Claims 122-126 and 129-131 are supported by a credible, specific and substantial asserted utility, and thus meet the utility requirement of 35 U.S.C. §§101/112, first paragraph

The sole basis for the Examiner's rejection of Claims 122-126 and 129-131 under this section is that the data presented in Example 170 of the present specification is allegedly insufficient under the present legal standards to establish a patentable utility under 35 U.S.C. §101 for the presently claimed subject matter.

Claims 122-126 and 129-131 stand further rejected under 35 U.S.C. §112, first paragraph, allegedly "since the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility for the reasons set forth above, one skilled in the art clearly would not know how to use the claimed invention."

Appellants strongly disagree and, therefore, respectfully traverse the rejection.

A. The Legal Standard For Utility Under 35 U.S.C. §101

According to 35 U.S.C. §101:

Whoever invents or discovers any new and *useful* process, machine, manufacture, or composition of matter, or any new and *useful* improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.
(Emphasis added).

In interpreting the utility requirement, in *Brenner v. Manson*,¹ the Supreme Court held that the *quid pro quo* contemplated by the U.S. Constitution between the public interest and the interest of the inventors required that a patent Appellant disclose a "substantial utility" for his or her invention, *i.e.*, a utility "where specific benefit exists in currently available form."² The Court concluded that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion. A patent system must be related to the world of commerce rather than the realm of philosophy."³

Later, in *Nelson v. Bowler*,⁴ the C.C.P.A. acknowledged that tests evidencing pharmacological activity of a compound may establish practical utility, even though they may not establish a specific therapeutic use. The Court held that "since it is crucial to provide researchers with an incentive to disclose pharmaceutical activities in as many compounds as possible, we conclude adequate proof of any such activity constitutes a showing of practical utility."⁵

In *Cross v. Iizuka*,⁶ the C.A.F.C. reaffirmed *Nelson*, and added that *in vitro* results might be sufficient to support practical utility, explaining that "*in vitro* testing, in general, is relatively less complex, less time consuming, and less expensive than *in vivo* testing. Moreover, *in vitro* results with the particular pharmacological activity are generally predictive of *in vivo* test results, *i.e.* there is a reasonable correlation there between."⁷ The Court perceived, "No insurmountable

¹ *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. (BNA) 689 (1966).

² *Id.* at 534, 148 U.S.P.Q. (BNA) at 695.

³ *Id.* at 536, 148 U.S.P.Q. (BNA) at 696.

⁴ *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. (BNA) 881 (C.C.P.A. 1980).

⁵ *Id.* at 856, 206 U.S.P.Q. (BNA) at 883.

⁶ *Cross v. Iizuka*, 753 F.2d 1047, 224 U.S.P.Q. (BNA) 739 (Fed. Cir. 1985).

⁷ *Id.* at 1050, 224 U.S.P.Q. (BNA) at 747.

difficulty" in finding that, under appropriate circumstances, "*in vitro* testing, may establish a practical utility."⁸

The case law has also clearly established that Appellants' statements of utility are usually sufficient, unless such statement of utility is unbelievable on its face.⁹ The PTO has the initial burden to prove that Appellants' claims of usefulness are not believable on their face.¹⁰ In general, an Appellant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. §101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope."^{11, 12}

Compliance with 35 U.S.C. §101 is a question of fact.¹³ The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration.¹⁴ Thus, to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the Appellant. The issue will then be decided on the totality of evidence.

⁸ *Id.*

⁹ *In re Gazave*, 379 F.2d 973, 154 U.S.P.Q. (BNA) 92 (C.C.P.A. 1967).

¹⁰ *Ibid.*

¹¹ *In re Langer*, 503 F.2d 1380,1391, 183 U.S.P.Q. (BNA) 288, 297 (C.C.P.A. 1974).

¹² *See also In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (C.C.P.A. 1980); *In re Irons*, 340 F.2d 974, 144 USPQ 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 USPQ 209, 212-13 (C.C.P.A. 1977).

¹³ *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. (BNA) 592, 596 (Fed. Cir. 1983) *cert. denied*, 469 US 835 (1984).

¹⁴ *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d (BNA) 1443, 1444 (Fed. Cir. 1992).

The well established case law is clearly reflected in the Utility Examination Guidelines (“Utility Guidelines”),¹⁵ which acknowledge that an invention complies with the utility requirement of 35 U.S.C. §101, if it has at least one asserted “specific, substantial, and credible utility” or a “well-established utility.” Under the Utility Guidelines, a utility is “specific” when it is particular to the subject matter claimed. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic without also identifying the conditions that are to be diagnosed.

In explaining the “substantial utility” standard, M.P.E.P. §2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a ‘substantial’ utility.”¹⁶ Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement,¹⁷ gives the following instruction to patent examiners: “If the Appellant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility.”

B. Proper Application of the Legal Standard

Appellants respectfully submit that the data presented in Example 170 starting on page 539 of the specification of the specification and the cumulative evidence of record, which underlies the current dispute, indeed support a “specific, substantial and credible” asserted utility for the presently claimed invention.

¹⁵ 66 Fed. Reg. 1092 (2001).

¹⁶ M.P.E.P. §2107.01.

¹⁷ M.P.E.P. §2107 II (B)(1).

Patentable utility for the PRO809 polypeptides is based upon the gene amplification data for the gene encoding the PRO809 polypeptide. Example 170 describes the results obtained using a very well-known and routinely employed polymerase chain reaction (PCR)-based assay, the TaqMan™ PCR assay, also referred to herein as the gene amplification assay. This assay allows one to quantitatively measure the level of gene amplification in a given sample, say, a tumor extract, or a cell line. It was well known in the art at the time the invention was made that gene amplification is an essential mechanism for oncogene activation. Appellants isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9 (pages 539 onwards of the specification), including primary lung and colon cancers of the type and stage indicated in Table 8 (page 546). The tumor samples were tested in triplicates with Taqman™ primers and with internal controls, beta-actin and GADPH in order to quantitatively compare DNA levels between samples (page 548, lines 33-34). As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control (page 539, lines 27-29) and also, no-template controls (page 548, lines 33-34). The results of TaqMan™ PCR are reported in ΔC_t units, as explained in the passage on page 539, lines 37-39. One unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, two units correspond to 4-fold, 3 units to 8-fold amplification and so on. Using this PCR-based assay, Appellants showed that the gene encoding for PRO809 was amplified, that is, it showed approximately 1.05-1.61 ΔC_t units for lung tumors which corresponds to $2^{1.05}$ - $2^{1.61}$ -fold amplification in lung, or 2.070 to 3.053 fold in five different lung primary tumors.

However, the Examiner states regarding the teachings of the Goddard Declaration that “the argument has been fully considered but is not deemed persuasive because the 2-3 fold amplification of the nucleic acid is not predictive of protein levels” (page 3, second paragraph of the Final Office Action mailed February 23, 2005). The Examiner further refers to Sen *et al.* (page 3 of the Final Office Action mailed February 23, 2005) to show that “cancerous tissues are known to be aneuploid” and that “pre-cancerous tissues *may* be aneuploid....Aneuploidy is also a feature of damaged tissue, and is commonly found in colon and lung tissues, which are subject to environmental damage. It does not invariably lead to cancer.”

Appellants respectfully point out that the Declaration by Dr. Audrey Goddard presented in their response mailed November 4, 2004 provides a statement by an expert in the relevant art

that “fold amplification” values of at least 2-fold are considered significant in the TaqMan™ PCR gene amplification assay. Appellants particularly draw the Board's attention to page 3 of the Goddard Declaration which clearly states that:

It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is **useful as a marker for the diagnosis of cancer**, for monitoring cancer development and/or for measuring the efficacy of cancer therapy. (Emphasis added).

Accordingly, the 2.070 to 3.053 fold in five different lung primary tumors would be considered significant and credible by one skilled in the art, based upon the facts disclosed in the Goddard Declaration.

Further, as any skilled artisan in the field of oncology would easily appreciate, not all tumor markers are generally associated with every tumor, or even, with most tumors. In fact, some tumor markers are useful for identifying rare malignancies. That is, the association of the tumor marker with a particular type of tumor lesion may be rare, or, the occurrence of that particular kind of tumor lesion itself may be rare. In either event, even these rare tumor markers, which may not give a positive hit with most common tumors, have great value in tumor diagnosis, and consequently, in tumor prognosis. The skilled artisan would know that such tumor markers are very useful for better classification of tumors. Therefore, whether the PRO809 gene is amplified in five lung in most tumors is not relevant to its identification as a tumor marker, or its patentable utility. Rather, whether the amplification data for PRO809 is significant is what lends support to its usefulness as a tumor marker. It was well known in the art at the time of filing of the application that gene amplification, which occurs in most solid tumors like lung cancers, is generally associated with poor prognosis. Therefore, the PRO809 gene becomes an important diagnostic marker to identify such malignant lung cancers, even if the malignancy associated with PRO809 molecule is a rare occurrence. Accordingly, the present specification clearly discloses enough evidence that the gene encoding the PRO809 polypeptide

is significantly amplified in certain types of lung tumors and is therefore, a valuable diagnostic marker for identifying certain types of lung cancers.

Regarding aneuploidy, Appellants had submitted the Ashkenazi Declaration to show that “detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy.” Regarding Sen, Appellants agree that aneuploidy can be a feature of damaged tissue as well, besides cancerous or pre-cancerous tissue, and may not invariably lead to cancer. However, Appellants maintain that Sen *et al.* support the Appellants’ position that even aneuploidy, which may be a feature of either cancerous or pre-cancerous tissue or damaged tissue, is still useful to diagnose a propensity for cancer or cancer itself. For instance, the art in lung cancer at the time of filing of the instant application shows that “epithelial tumors develop through a multistep process driven by genetic instability” in damaged lesions. Many articles published around June 23, 1999 (the effective filing date of this application) studied damaged or premalignant lesions and suggested that epithelial tumors develop through a multistep process driven by genetic instability and that a subset of the same molecular changes found in associated tumors were also found in premalignant lesions, suggesting that these premalignant lesions might represent precursor lesions for associated tumors, *i.e.*, a manifestation of a multistep tumorigenesis process. Based on the well-known art, Appellants submit that there is utility in identifying genetic biomarkers in epithelial tissues at cancer risk.

Taken together, even if the observed PRO809 gene amplification were due to chromosomal aneuploidy (which Appellants do not concede), such an observation would still support at least one utility for the PRO809 gene because it helps in identifying individuals at significantly increased cancer risk. Therefore, the art supports at least one utility for the PRO809 gene, that is, as a genetic biomarker for cancer or precancerous cells or damaged tissue. As one skilled in the art would clearly know, early detection of lung cancer provides information in advance about risk assessment, prognosis and therapy for lung cancer. Accordingly, the instant polynucleotides find utility as a diagnostic for cancer or for individuals at risk for developing lung cancer.

C. **A prima facie case of lack of utility has not been established**

The Examiner further cited Hu *et al.*, to show that “the literature cautions researchers against drawing conclusions based on small changes in transcript expression levels between normal and cancerous tissues” (Page 5 of the Final Office Action mailed February 23, 2005).

First of all, as discussed above, the increase in DNA copy number for the PRO809 gene is significant. Further, Appellants respectfully submit that, contrary to the Examiner’s assertion, the cited Hu *et al.* reference does not conclusively establish a *prima facie* case for lack of utility for the PRO809 molecule. The Hu *et al.* reference is entitled “Analysis of Genomic and Proteomic Data using Advanced Literature Mining” (emphasis added). Therefore, as the title itself suggests, the conclusions in this reference are based upon statistical analysis of information obtained from published literature, and not from experimental data. Hu *et al.* performed statistical analysis to provide evidence for a relationship between mRNA expression and biological function of a given molecule (as in disease). The conclusions of Hu *et al.* however, only apply to a specific type of breast tumor (estrogen receptor (ER)-positive breast tumor) and cannot be generalized to breast cancer genes in general, let alone to cancer genes in general. Interestingly, the observed correlation was only found among ER-positive (breast) tumors not ER-negative tumors.” (See page 412, left column).

Moreover, the analytical methods utilized by Hu *et al.* have certain statistical drawbacks, as the authors themselves admit. For instance, according to Hu *et al.*, “*different statistical methods*” were applied to “*estimate the strength of gene-disease relationships and evaluated the results.*” (See page 406, left column, emphasis added). Using these different statistical methods, Hu *et al.* “[a]ssessed the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation.” (See page 411, left column). As is well known in the art, different statistical methods allow different variables to be manipulated to affect the resulting outcome. In this regard, the authors disclose that, “Initial attempts to search the literature” using the list of genes, gene names, gene symbols, and frequently used synonyms generated by the authors “revealed several sources of false positives and false negatives.” (See page 406, right column). The authors add that the false positives caused by “duplicative and unrelated meanings for the term” were “difficult to manage.” Therefore, in order to minimize such false positives, Hu *et al.* disclose that these terms “had to be eliminated entirely, thereby reducing the false

positive rate but unavoidably under-representing some genes.” *Id.* (emphasis added). Hence, Hu *et al.* had to manipulate certain aspects of the input data, in order to generate, in their opinion, meaningful results. Further, because the frequency of citation for a given molecule and its relationship to disease only reflects the current research interest of a molecule, and not the true biological function of the molecule, as the authors themselves acknowledge, the “[r]elationship established by frequency of co-citation do not necessarily represent a true biological link.” (See page 411, right column). Therefore, based on these findings, the authors add, “[t]his may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently.” *Id.* (Emphasis added). In other words, some molecules may have been underrepresented merely because they were less frequently cited or studied in literature compared to other more well-cited or studied genes. Therefore, Hu *et al.*’s conclusions are not based on genes/mRNA *in general*.

Therefore, Appellants submit that, based on the nature of the statistical analysis performed herein, and in particular, based on Hu’s analysis of *one* class of genes, namely, the estrogen receptor (ER)-positive breast tumor genes, the conclusions drawn by the Examiner, namely that, “genes displaying a 5-fold change or less (mRNA expression) in tumors compared to normal showed no evidence of a correlation between altered gene expression and a known role in the disease (in general)” is not reliably supported.

Therefore, when the proper legal standard is used, a *prima facie* case of lack of utility has not been met based on the cited reference Hu *et al.* by the Examiner.

On the contrary, Appellants submit that Example 170 in the specification further discloses that, “(a)mplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as colon, lung, breast and other cancers and diagnostic determination of the presence of those cancers” (Emphasis added). Besides, Appellants have submitted ample evidence (discussed below) to show that, in general, if a gene is amplified in cancer, it is “more likely than not” likely that the encoded protein will also be expressed at an elevated level.

For support, Appellants presented the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* (made of record in Appellants’ Response filed November 4, 2004), who collectively teach

that in general, for most genes, DNA amplification increases mRNA expression. The results presented by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* are based upon wide ranging analyses of a large number of tumor associated genes. Orntoft *et al.* studied transcript levels of 5600 genes in malignant bladder cancers, many of which were linked to the gain or loss of chromosomal material, and found that in general (18 of 23 cases) chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Hyman *et al.* compared DNA copy numbers and mRNA expression of over 12,000 genes in breast cancer tumors and cell lines, and found that there was evidence of a prominent global influence of copy number changes on gene expression levels. In Pollack *et al.*, the authors profiled DNA copy number alteration across 6,691 mapped human genes in 44 predominantly advanced primary breast tumors and 10 breast cancer cell lines, and found that on average, a 2-fold change in DNA copy number was associated with a corresponding 1.5-fold change in mRNA levels. In summary, the evidence supports the Appellants' position that gene amplification is more likely than not predictive of increased mRNA and polypeptide levels.

Second, the Declaration of Dr. Paul Polakis (made of record in Appellants' Response filed November 4, 2004), principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, explains that in the course of Dr. Polakis' research using microarray analysis, he and his co-workers identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. Appellants submit that Dr. Polakis' Declaration was presented to support the position that there is a correlation between mRNA levels and polypeptide levels, the correlation between gene amplification and mRNA levels having already been established by the data shown in the Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* articles. Appellants further emphasize that the opinions expressed in the Polakis Declaration, including in the above quoted statement, are all based on factual findings. For instance, antibodies binding to about 30 of these tumor antigens were prepared, and mRNA and protein levels were compared. In approximately 80% of the cases, the researchers found that increases in the level of a particular mRNA correlated with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells. Therefore, Dr. Polakis' research, which is referenced in his

Declaration, shows that, in general, there is a correlation between increased mRNA and polypeptide levels.

Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip® arrays in 2004. Clearly, the research community believe that the information obtained from these chips is useful (*i.e.*, that it is more likely than not that the results are informative of protein levels).

However, in the Final Office Action (page 4) mailed February 23, 2005, the Examiner alleges regarding Orntoft *et al.*, Hyman *et al.* and Pollack *et al.* that:

“Orntoft *et al.* could only compare the levels of about 40 well-resolved and focused abundant proteins (see abstract). It would appear that Applicants have provided no fact or evidence concerning a correlation between such low levels of amplification of DNA, found only in a minority tested tumors which were not characterized on the basis of those in the Orntoft publication, and an associated rise in level of the encoded protein”.

“Hyman *et al.* found 44% of highly amplified genes showing overexpression at the mRNA level, and 10.5% of highly overexpressed genes being amplified; thus, even at the level of high amplification and high overexpression, the two do not correlate.....Hyman does not examine protein expression”

“Similarly, Pollack *et al.*, does not analyze protein levels, nor does Pollack support the assertion that it is predictable, on the basis of the minimal increase in copy number of SEQ ID NO:222 that the protein would accordingly be found at altered levels. Accordingly, **it remains that the significance of the gene amplification data is questionable, and cannot be predictably extrapolated as applying to the claimed protein**” (Emphasis added).

Appellants disagree on all of the above points. Appellants respectfully submit that the Examiner has misrepresented the data discussed in the above articles as explained below. For example, contrary to what the Examiner says, in Orntoft *et al.*, 1,800 genes that yielded an increase or decrease in mRNA expression in two invasive tumors compared to the two non-invasive papillomas were then mapped to chromosomal locations. The chromosomes had already been analyzed for amplification by hybridizing tumor DNA to normal metaphase chromosomes (CGH). Orntoft *et al.* used CGH alterations as the independent variable and estimated the frequency of expression alterations of the 1,800 genes in the chromosomal areas.

Orntoft *et al.* found that in general (77% and 80% concordance) areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression (see page 40). Orntoft *et al.* state, "For both tumors TCC733 ($p < 0.015$) and TCC827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology" (see page 41, column 1). Orntoft *et al.*, also studied the relation between altered mRNA and protein levels using 2D-PAGE analysis. Orntoft *et al.* state, "In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations.... 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays." (See page 42, column 2 to page 34, column 2). Accordingly, Orntoft *et al.* clearly support Appellants' position that proteins expressed by genes that are amplified in tumors are useful as cancer markers. Orntoft *et al.* did not limit their findings to only those regions of amplified gene clusters.

Appellants also respectfully submit that the Examiner has mischaracterized the studies done by Hyman *et al.* and Pollack *et al.* and their analyses. Hyman *et al.* and Pollack *et al.* did gene-by-gene analysis across all chromosomes. In Hyman *et al.*, 13,824 cDNA clones were placed on glass slides in a microarray and genomic DNA from breast cancer cell lines and normal human WBCs was hybridized to the cDNA sequences. For expression analysis, RNA from tumor cell lines was hybridized on the same microarrays. The 13,824 arrayed cDNA clones were analyzed for gene expression and gene copy number in 14 breast cancer cell lines. Hyman *et al.* state, "The results illustrate a considerable influence of copy number on gene expression patterns." For example, Hyman *et al.* teach that "[u]p to 44% of the highly amplified transcripts (CGH ratio, > 2.5) were overexpressed (*i.e.*, belonged to the global upper 7% of expression ratios) compared with only 6% for genes with normal copy number." (See page 6242, column 1). Further, Hyman *et al.* state that "[t]he cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome." (See page 6242, column 2). Therefore, the analysis performed by Hyman *et al.* was on a gene-by gene basis, and clearly shows that "it is more likely than not" that a gene which is amplified in tumor cells will have increased gene expression.

In Pollack *et al.*, DNA copy number alteration across 6,691 mapped human genes in 44 predominantly advanced primary breast tumors and 10 breast cancer cell lines was profiled. Pollack *et al.* further state, "Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells." (See Abstract). "Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4, and representing 91 different genes), 62% (representing 54 different genes; ...) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4)." (See page 12966, column 1). Therefore, the analysis performed by Pollack *et al.* was also done on a gene-by gene basis, and clearly shows that "it is more likely than not" that a gene which is amplified in tumor cells will have increased gene expression.

Thus, the Examiner appears to disregard the ample evidence provided in the above referenced articles based on misinterpretations of their teachings. Appellants submit that in fact, these articles lend significant support that for an amplified gene, it is more likely than not that the protein will also be overexpressed and would be viewed as reasonable and credible by one of ordinary skill in the art. The "more likely than not" standard is a much lower standard than a "necessary" correlation or "accurate" prediction, and is clearly met in the claimed invention. Moreover, the Examiner has not cited any evidence or advanced any arguments as to why Appellants' statement of overexpression of protein would not be credible. Accordingly, this point is believed to be moot.

Taken together, all of the submitted evidence supports the Appellants' position that, in the majority of amplified genes, increased gene amplification levels, more likely than not, predict increased mRNA and polypeptide levels, which clearly meets the utility standards described above. Hence, one of skill in the art would reasonably expect that, based on the gene amplification data of the PRO809 gene, the PRO809 polypeptide is concomitantly overexpressed in the lung tumors studied as well.

The Examiner also finds the declaratory evidence presented by Polakis in this case, for what Appellants consider legally inappropriate reasons, "non-persuasive" (page 5 of Final Office action mailed February 23, 2005). The Examiner further states that the Ashkenazi Declaration is

not persuasive because in the instant case “there is no evidence as to whether the gene products (such as the claimed polypeptide) are overexpressed or not. Further research is required to determine such” (page 6, first paragraph, lines 6-7). Appellants respectfully disagree.

First of all, Appellants respectfully draw the Examiner’s attention to case law that clearly establishes that in considering affidavit evidence, the Examiner must consider all of the evidence of record anew (*In re Rinehart*, 531 F.2d 1084, 189 USPQ 143 (C.C.P.A. 1976) and *In re Piasecki*, 745 F.2d. 1015, 226 USPQ 881 (Fed. Cir. 1985)). “After evidence or argument is submitted by the applicant in response, patentability is determined on the totality of the record, by a preponderance of the evidence with due consideration to persuasiveness of argument” (*In re Alton*, 37 USPQ2d 1578 (Fed. Cir 1966) at 1584 quoting *In re Oetiker*, 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992)). Furthermore, the Federal Court of Appeals held in *In re Alton*, “We are aware of no reason why opinion evidence relating to a fact issue should not be considered by an examiner” (*In re Alton, supra.*). Appellants further draw the Examiner's attention to the Utility Examination Guidelines (Part IIB, 66 Fed. Reg. 1098 (2001)) which states,

“Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered.”

The statement in question from the Polakis Declaration that “it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell” is based on his own experimental findings, which is clearly set forth in the Declaration. Accordingly, the fact-based conclusions of Dr. Polakis and Dr. Ashkenazi would be considered reasonable and accurate by one skilled in the art. Thus, barring evidence to the contrary, Appellants maintain that the fold amplification disclosed for the PRO809 gene is significant and forms the basis for the utility claimed for the PRO809 polypeptide herein.

Appellants further submit that even if there were no correlation between gene amplification and increased mRNA/protein expression, (which Appellants expressly do not concede), a polypeptide encoded by an amplified gene in cancer would **still** have a specific,

substantial, and credible utility as explained below. As the Declaration of Dr. Avi Ashkenazi (submitted with Appellants' Response filed November 4, 2004) explains:

"even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment."

Thus, even if over-expression of the gene product does not parallel gene amplification in certain tumor types, parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified in a tumor, but the corresponding gene product is not over-expressed, the clinician will decide not to treat a patient with agents that target that gene product. This not only saves money, but also has the benefit that the patient can avoid exposure to the side effects associated with such agents.

This utility is further supported by the teachings of the article by Hanna and Mornin. (Pathology Associates Medical Laboratories, August (1999), submitted with the Response filed November 4, 2004). The article teaches that the HER-2/neu gene has been shown to be amplified and/or over-expressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinomas. Further, the article teaches that diagnosis of breast cancer includes testing both the amplification of the HER-2/neu gene (by FISH) as well as the over-expression of the HER-2/neu gene product (by IHC). Even when the protein is not over-expressed, the assay relying on both tests leads to a more accurate classification of the cancer and a more effective treatment of it.

The Examiner asserts that,

"Hanna *et al.* supports the instant rejection, in that Hanna *et al.* show that gene amplification does not reliably correlate with polypeptide overexpression, and thus the level of polypeptide expression must be tested empirically." (Page 6 of the Final Office Action mailed February 23, 2005).

Appellants respectfully point out that the Examiner appears to have misread Hanna *et al.* Hanna *et al.* clearly state that gene amplification (as measured by FISH) and polypeptide expression (as measured by immunohistochemistry, IHC) are well correlated ("in general, FISH

and IHC results correlate well" (Hanna *et al.* p. 1, col. 2)). It is only a subset of tumors which show discordant results. Thus, Hanna *et al.* support Appellants' position rather well that it is more likely than not that gene amplification correlates with increased polypeptide expression. The Examiner appears to view such testing described in the Ashkenazi Declaration and the Hanna paper as experiments involving further characterization of the PRO809 polypeptide itself. On the contrary, such testing is for the purpose of characterizing not the PRO809 polypeptide, but the tumors in which the gene encoding PRO809 is amplified. That is, such further testing or research is for the purpose of characterizing the tumors into medically relevant categories in which the gene encoding PRO809 is/is not amplified, and such techniques were routine in the art of clinical oncology at the time of filing of the instant application, as evidenced by the teaching of Hanna and Mornin.

Thus, based on the asserted utility for PRO809 in the diagnosis of selected lung tumors, the reduction to practice of the instantly claimed protein sequence of SEQ ID NO:223 in the present application, the disclosure of the step-by-step protocols for making chimeric PRO polypeptides, including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin in the specification (at page 374, lines 24 to page 375, line 9), the disclosure of a step-by-step protocol for making and expressing PRO809 in appropriate host cells (in Examples 140-143 and page 376, line 12), the step-by-step protocol for the preparation, isolation and detection of monoclonal, polyclonal and other types of antibodies against the PRO809 protein in the specification (at pages 390-395) and the disclosure of the gene amplification assay in Example 170, the skilled artisan would know exactly how to make and use the claimed polypeptide for the diagnosis of lung cancers. Appellants submit that based on the detailed information presented in the specification and the advanced state of the art in oncology, the skilled artisan would have found such testing routine and not 'undue.'

Therefore, since the instantly claimed invention is supported by either a credible, specific and substantial asserted utility or a well-established utility, and since the present specification clearly teaches one skilled in the art "how to make and use" the claimed invention without undue experimentation, Appellants respectfully request reconsideration and reversal of this outstanding rejections under 35 U.S.C. §101 and §112, First Paragraph to Claims 122-126 and 129-131.

ISSUE 2: Claims 122-124 and 130-131 satisfy the written description requirement of 35 U.S.C. §112, First Paragraph

Claims 122-124 and 130-131 stand rejected under 35 U.S.C. §112, first paragraph, as allegedly containing "subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention." In particular, the Examiner has asserted that "All applicants have presented is a single nucleic acid found to be slightly amplified in a small proportion of cancers, and the germ of an idea that there might be variants of the nucleic acid that would be similarly associated. There is no evidence of the actual conception of such nucleic acids, nor is there any evidence of record that they exist" (Page 7 of the Final Office Action mailed February 23, 2005).

Appellants respectfully disagree.

A. The Legal Test for Written Description

The well-established test for sufficiency of support under the written description requirement of 35 U.S.C. §112, first paragraph is "whether the disclosure of the application as originally filed reasonably conveys to the artisan that the inventor had possession at that time of the later claimed subject matter, rather than the presence or absence of literal support in the specification for the claim language."^{18,19} The adequacy of written description support is a factual issue and is to be determined on a case-by-case basis.²⁰ The factual determination in a written description analysis depends on the nature of the invention and the amount of knowledge imparted to those skilled in the art by the disclosure.^{21, 22}

¹⁸ *In re Kaslow*, 707 F.2d 1366, 1374, 212 USPQ 1089, 1096 (Fed. Cir. 1983).

¹⁹ *See also Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1563, 19 USPQ2d at 1116 (Fed. Cir. 1991).

²⁰ *See e.g., Vas-Cath*, 935 F.2d at 1563; 19 USPQ2d at 1116.

²¹ *Union Oil v. Atlantic Richfield Co.*, 208 F.2d 989, 996 (Fed. Cir. 2000).

²² *See also* M.P.E.P. §2163 II(A).

In *Environmental Designs, Ltd. v. Union Oil Co.*,²³ the Federal Circuit held, "Factors that may be considered in determining level of ordinary skill in the art include: (1) the educational level of the inventor; (2) type of problems encountered in the art; (3) prior art solutions to those problems; (4) rapidity with which innovations are made; (5) sophistication of the technology; and (6) educational level of active workers in the field."²⁴ Further, the "hypothetical 'person having ordinary skill in the art' to which the claimed subject matter pertains would, of necessity have the capability of understanding the scientific and engineering principles applicable to the pertinent art."^{25, 26}

B. The Disclosure Provides Sufficient Written Description for the Claimed Invention

Appellants respectfully submit that the instant specification evidences the actual reduction to practice of the amino acid sequence of SEQ ID NO: 223. Appellants also submit that the specification provides ample written support for determining percent sequence identity between two amino acid sequences (See pages 306-308, line 14 onwards). In fact, the specification teaches specific parameters to be associated with the term "percent identity" as applied to the present invention. The specification further provides detailed guidance as to changes that may be made to a PRO polypeptide without adversely affecting its activity (page 372, line 36 to page 373, line 17). This guidance includes a listing of exemplary and preferred substitutions for each of the twenty naturally occurring amino acids (Table 6, page 372). Accordingly, one of skill in the art could identify whether the variant PRO809 sequence falls within the parameters of the claimed invention. Once such an amino acid sequence was identified, the specification sets forth methods for making the amino acid

²³ 713 F.2d 693, 696, 218 USPQ 865, 868 (Fed. Cir. 1983), *cert. denied*, 464 U.S. 1043 (1984).

²⁴ *See also* M.P.E.P. §2141.03.

²⁵ *Ex parte Hiyamizu*, 10 USPQ2d 1393, 1394 (Bd. Pat. App. & Inter. 1988) (emphasis added).

²⁶ *See also* M.P.E.P. §2141.03.

sequences (see page 376, line 9) and methods of preparing the PRO polypeptides (see Examples 140-143).

Currently pending Claims 122-126 recite the functional limitation that the nucleic acid encoding the claimed polypeptides are amplified in lung tumors. Appellants further submit that the specification provides ample written support for detecting and quantifying amplification of such nucleic acids in several tumors and/or cell lines as described in Example 170. Example 170 of the present application provides step-by-step guidelines and protocols for the gene amplification assay. By following this disclosure, one skilled in the art would know that it is easy to test whether a gene encoding a variant PRO809 protein is amplified in lung tumors by the methods set forth in Example 170.

More recently, in *Enzo Biochem., Inc. v. Genprobe, Inc.* 296 F.3d 1316 (Fed. Cir. 2002), the court adopted the standard that "the written description requirement can be met by 'showing that the invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics, . . . i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.'" *Id.* at 1324. While the invention in *Enzo* was still a DNA, the holding has been treated as being applicable to proteins as well. Indeed, the court adopted the standard from the USPTO's Written Description Examination Guidelines, which apply to both proteins and nucleic acids.

Accordingly, current applicable case law holds that biological sequences are not adequately described solely by a description of their desired functional activities. The instant claims meet the standard set by the *Enzo* court in that the claimed sequences are defined not only by functional properties, but also by structural limitations. It is well established that a combination of functional and structural features may suffice to describe a claimed genus. "An applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics which provide evidence that applicant was in possession of the claimed invention, i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation

between function and structure, or some combination of such characteristics."²⁷ Thus, the genus of polypeptides with at least 95-99% sequence identity to SEQ ID NO:223, which possess the functional property of having a nucleic acid which is amplified in lung tumor would meet the requirement of 35 U.S.C. §112, first paragraph, as providing adequate written description. Accordingly, one skilled in the art would have known that Appellants had knowledge and possessed the claimed polypeptides with 95-99% sequence identity to SEQ ID NO: 223 whose encoding nucleic acids were amplified in lung tumors. The recited property of amplification of the encoding gene adds to the characterization of the claimed polypeptide sequences in a manner that one of skill in the art could readily assess and understand.

As discussed above, Appellants have recited structural features, namely, 95% sequence identity to SEQ ID NO: 223, which are common to the genus. Appellants have also provided guidance as to how to make the recited variants of SEQ ID NO:223, including listings of exemplary and preferred sequence substitutions. The genus of claimed polypeptides is further defined by having a specific functional activity for the encoding nucleic acids. Accordingly, a description of the claimed genus has been achieved.

For the above reasons, the specification provides adequate written description for polypeptides having at least 95% identity to SEQ ID NO:223 wherein the nucleic acid encoding the polypeptide is amplified in lung tumor. Accordingly, Appellants respectfully request reconsideration and reversal of the written description rejection of Claims 119-126 under 35 U.S.C. §112, first paragraph.

ISSUE 3. Claims 122-126, 129-131 and 135-138 are patentable over Hillier *et al.*, 1995 (EST clone H74302) in view of Sibson *et al.*, WO94/01548

Claims 122-126 and 129-131 remain rejected under 35 U.S.C. §103(a) as being unpatentable over clone H74302 (EST isolated by Hillier *et al.*, 1995) in view of Sibson *et al.*, WO94/01548.

The instant application claims **PRO809 polypeptides**. Appellants submit that the Hiller EST clone does not disclose nor reduces to practice the sequence of SEQ ID NO: 223. Further,

²⁷ M.P.E.P. §2163 II(A)(3)(a)

Hillier does not disclose that their EST was part of a cDNA sequence that coded for a polypeptide (see enclosed alignment Item 13, Evidence Appendix). Thus, Hillier *et al.* did not teach nor reduce to practice the instantly claimed polypeptide sequences. Further, one of skill in the art would not have been able to envision the cDNA clone defined in SEQ ID NO:222 from the teachings of Hillier alone because the Hillier sequence lacks the C-terminus of the cDNA and therefore, Hillier *et al.* does not anticipate the instant invention. Further, the secondary reference Sibson *et al.* also do not teach or anticipate the instant invention.

Accordingly, this rejection under 35 U.S.C. §103(a) should be withdrawn.

CONCLUSION

For the reasons given above, Appellants submit that present specification clearly describes, details and provides a patentable utility for the claimed invention. Moreover, it is respectfully submitted that based upon this disclosed patentable utility, the present specification clearly teaches "how to use" the presently claimed polypeptide. As such, Appellants respectfully request reconsideration and reversal of the outstanding rejection of Claims 122-126 and 129-131.

The Commissioner is authorized to charge any fees which may be required, including extension fees, or credit any overpayment to Deposit Account No. 08-1641 (referencing Attorney's Docket No. 39780-2730 P1C15).

Respectfully submitted,

Date: December 22, 2005

By: Lohi M. Jui, Reg. No. 37,047
for Daphne Reddy (Reg. No. 53,507)

HELLER EHRMAN LLP
275 Middlefield Road
Menlo Park, California 94025-3506
Telephone: (650) 324-7000
Facsimile: (650) 324-0638

IX. CLAIMS APPENDIX

Claims on Appeal

122. An isolated polypeptide having at least 95% amino acid sequence identity to:
- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 223;
 - (b) the amino acid sequence of the polypeptide of SEQ ID NO: 223, lacking its associated signal peptide; or
 - (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025, wherein the nucleic acid encoding said polypeptide is amplified in lung tumors.
123. An isolated polypeptide having at least 99% amino acid sequence identity to:
- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 223;
 - (b) the amino acid sequence of the polypeptide of SEQ ID NO: 223, lacking its associated signal peptide; or
 - (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025, wherein the nucleic acid encoding said polypeptide is amplified in lung tumors.
124. An isolated polypeptide comprising:
- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 223;
 - (b) the amino acid sequence of the polypeptide of SEQ ID NO: 223, lacking its associated signal peptide;
 - (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025, wherein the nucleic acid encoding said polypeptide is amplified in lung tumors.
125. The isolated polypeptide of Claim 124 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 223.
126. The isolated polypeptide of Claim 124 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 223, lacking its associated signal peptide.

129. The isolated polypeptide of Claim 124 comprising the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025.

130. A chimeric polypeptide comprising a polypeptide according to Claim 124 fused to a heterologous polypeptide.

131. The chimeric polypeptide of Claim 130, wherein said heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin.

X. EVIDENCE APPENDIX

1. Declaration of Paul Polakis, Ph.D. under 35 C.F.R. §1.132.
2. Declaration of Avi Ashkenazi, Ph.D. under 35 C.F.R. §1.132, with attached Exhibit A (Curriculum Vitae).
3. Declaration of Audrey Goddard, Ph.D. under 35 C.F.R. §1.132, with attached Exhibits A-G:
 - A. Curriculum Vitae of Audrey D. Goddard, Ph.D.
 - B. Higuchi, R. et al., "Simultaneous amplification and detection of specific DNA sequences," *Biotechnology* 10:413-417 (1992).
 - C. Livak, K.J., et al., "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *PCR Methods Appl.* 4:357-362 (1995).
 - D. Heid, C.A. et al., "Real time quantitative PCR," *Genome Res.* 6:986-994 (1996).
 - E. Pennica, D. et al., "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors," *Proc. Natl. Acad. Sci. USA* 95:14717-14722 (1998).
 - F. Pitti, R.M. et al., "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer," *Nature* 396:699-703 (1998).
 - G. Bieche, I. et al., "Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer," *Int. J. Cancer* 78:661-666 (1998).
4. Orntoft, T.F., et al., "Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-Invasive and Invasive Human Transitional Cell Carcinomas," *Molecular & Cellular Proteomics* 1:37-45 (2002).
5. Hyman, E., et al., "Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer," *Cancer Research* 62:6240-6245 (2002).
6. Pollack, J.R., et al., "Microarray Analysis Reveals a Major Direct Role of DNA Copy Number Alteration in the Transcriptional Program of Human Breast Tumors," *Proc. Natl. Acad. Sci. USA* 99:12963-12968 (2002).
7. Hanna et al., "HER-2/neu Breast Cancer Predictive Testing," Pathology Associates Medical Laboratories (1999).

8. Pennica *et al.*, "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors," *Proc. Natl. Acad. Sci. USA* **95**:14717-14722 (1998).

9. Sen S., "Aneuploidy and Cancer", *Current Opinion in Oncology*, **12**: 82-88, (2000).

10. Hillier L., Loci H74303, H73373, H58326, R02548, The WashU-Merck EST Project, (1995).

11. Hu *et al.*, "Analysis of genomic and proteomic data using advanced literature mining," *Journal of Proteome Research* **2**:405-412 (2003).

12. Sibson *et al.*, WO94/01548 (January 1994).

13. Alignment of the instantly claimed DNA 57836-1338 with EST sequence of Hillier *et al.*, (H74302).

Items 1-7 were submitted with Appellants' Response filed November 4, 2004, and were considered by the Examiner as indicated in the Final Office action mailed February 23, 2005.

Items 8-10 and 12 were made of record by the Examiner in the Office Action mailed June 3, 2004.

Item 11 was made of record by the Examiner in the Final Office Action mailed February 23, 2005.

Item 13 is submitted herewith in response to the Examiner's request for an alignment of the claimed nucleic acids with the sequence of the EST clone (see page 8 of Final Office Action mailed February 23, 2005).

XI. RELATED PROCEEDINGS APPENDIX

None.



DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

| | |
|--------------|---|
| 2002-present | Staff Scientist, Genentech, Inc S. San Francisco, CA |
| 1999- 2002 | Senior Scientist, Genentech, Inc., S. San Francisco, CA |
| 1997 -1999 | Research Director Onyx Pharmaceuticals, Richmond, CA |
| 1992- 1996 | Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA |
| 1991-1992 | Senior Scientist, Chiron Corporation, Emeryville, CA. |
| 1989-1991 | Scientist, Cetus Corporation, Emeryville CA. |
| 1987-1989 | Postdoctoral Research Associate, Genentech, Inc., South San Francisco, CA. |
| 1985-1987 | Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC |

1984-1985

Assistant Professor, Department of Chemistry,
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of
Biochemistry, Michigan State University
East Lansing, Michigan

PUBLICATIONS:

1. **Polakis, P. G.** and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.
2. **Polakis, P.G.** and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.
3. **Polakis, P. G.** and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.
4. Uhing, R.J., **Polakis, P.G.** and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.
5. **Polakis, P.G.**, Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.
6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.
7. **Polakis, P.G.**, Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.
8. **Polakis, P. G.**, Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Commun.** 160, 25-32.
9. **Polakis, P.**, Weber, R.F., Nevins, B., Didsbury, J. Evans, T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.
10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K⁺ Channels. **Cell.** 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.

14. **Polakis, P.G.** Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21^{ras} Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21^{rap1}. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21^{ras}-dependent Inhibition of Muscarinic Atrial K⁺ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P.** 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60^{v-src} Can Associate with and Phosphorylate the p21^{ras} GTPase Activating Protein. **Oncogene** 7,389-394.

22. **Polakis P.** and McCormick, F. 1992 Interactions Between p21^{ras} Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.
24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34^{cdc2} Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.
25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21^{rap1A} and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.
26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell. Biol.** 12, 4634-4642.
27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.
28. Janoueix-Lerosey, I., Polakis, P., Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.
29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.
30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21^{ras} with GAP, exchange factors, and its biological effector target. **J. Biol. Chem.** 268, 9157-9160.
31. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.
32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F., Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.
33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.
34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., Polakis, P. and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975
35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31
37. **Polakis, P.** 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α -catenin, β -catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β -catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.
40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Iol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3 β to the APC- β -catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β -catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. **Polakis, P.** The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P.** 1997 Loss of β -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and **Polakis, P.** 1997 Induction of a β -catenin-LEF-1 complex by wnt-1 and transforming mutants of β -catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β -catenin and glycogen synthase kinase 3 β . **Current Biology** 8, 573-581.

56. **Polakis, P.** 1998 The oncogenic activation of β -catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and **Paul Polakis.** 1999 The F-box protein β -TrCP associates with phosphorylated β -catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β -catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P.**, Casey PJ. 1999 Functional interaction between G α (z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β -catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.
- 61 **Polakis P**, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32
- 62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53
64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.
65. Peifer M, **Polakis P**. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.
66. **Polakis P**. 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.
67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.
68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.
69. Rubinfeld B, Tice DA, **Polakis P**. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.
70. **Polakis P**. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.
71. Tice DA, Soloviev I, **Polakis P**. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem.** 277, 6118-6123.
72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.

73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIs--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics**, April 8



THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.
App. No. : 09/903,925
Filed : July 11, 2001
For : SECRETED AND
TRANSMEMBRANE
POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan[®] PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi
Avi Ashkenazi, Ph.D.

Date: 9/15/03

CURRICULUM VITAE

Avi Ashkenazi

July 2003

Personal:

Date of birth: 29 November, 1956
Address: 1456 Tarrytown Street, San Mateo, CA 94402
Phone: (650) 578-9199 (home); (650) 225-1853 (office)
Fax: (650) 225-6443 (office)
Email: aa@gene.com

Education:

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel
1986: Ph.D. in Biochemistry, Hebrew University, Israel

Employment:

1983-1986: Teaching assistant, undergraduate level course in Biochemistry
1985-1986: Teaching assistant, graduate level course on Signal Transduction
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,
with D. Capon
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.
1994 -1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,
Genentech, Inc.
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.
1999 -2002: Staff Scientist in Molecular Oncology, Genentech, Inc.
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

Awards:

1988: First prize, The Boehringer Ingelheim Award

Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* 34, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* 118, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* 50, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* 14, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* 210, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotrophic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* 121, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* 236, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* 6, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* 238, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* **37**, 119-129 (1988).
11. Peralta, E. Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* **334**, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* **56**, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* **340**, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* **245**, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* **87**, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* **29**, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* **88**, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* **88**, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine_{1B} receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT_{1B} receptors. *Nature* **360**, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon γ by an interferon γ receptor immunoadhesin. *Immunology* **79**, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).
26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3⁺ effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by interferon γ and correction by interferon γ receptor immunoadhesin. *Blood* **83**, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* **5**, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT₂ receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- γ receptor β chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA.* **92**, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal E. coli bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- κ B and AP-1 transcription factors. *J. Biol. Chem.* 272, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 277, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* 7, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* 427, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* 160, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* 162, 5616-5623 (1999).
49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* 9, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN- α receptor 2. *J. Immunol.* **163**, 766-773 (1999).
52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563-571 (1999).
55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., *Ashkenazi, A., and *Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* 10, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* 29, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D. Mounho, B., Hillan, K., Totpal, K. DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* 7, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* 166, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* 7, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., *Ashkenazi, A., and *Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

Review articles:

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzymology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook (www.apnet.com/cytokinereference/).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

Book:

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

Talks:

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Halflife Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- γ signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

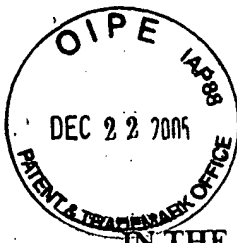
9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philadelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14th Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Switzerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

Issued Patents:

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).



PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | |
|---|---|
| In re Application of: Ashkenazi et al. | Group Art Unit: 1647 |
| Serial No.: 09/903,925 | Examiner: Fozia Hamid |
| Filed: July 11, 2001 | CERTIFICATE OF MAILING I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231 on Date _____ |
| For: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS | |

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.
2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

Serial No.: *

Filed: *

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl. 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

Serial No.: *

Filed: *

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003

Date

Audrey D. Goddard

Audrey D. Goddard, Ph.D.

AUDREY D. GODDARD, Ph.D.

Genentech, Inc.
1-DNA Way
South San Francisco, CA, 94080
650.225.6429
goddarda@gene.com

110 Congo St.
San Francisco, CA, 94131
415.841.9154
415.819.2247 (mobile)
agoddard@pacbell.net

PROFESSIONAL EXPERIENCE

Genentech, Inc.
South San Francisco, CA

1993-present

2001 - present Senior Clinical Scientist
Experimental Medicine / BioOncology, Medical Affairs

Responsibilities:

- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

Interests:

- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

1998 - 2001 Senior Scientist
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities:

- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a \$750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

Research:

- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

1993 - 1998 Scientist

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities

- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

Research:

- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with Idiopathic short stature*

**Imperial Cancer Research Fund
London, UK with Dr. Ellen Solomon**

1989-1992

6/89 –12/92 Postdoctoral Fellow

- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

**McMaster University
Hamilton, Ontario, Canada with Dr. G. D. Sweeney**

1983

5/83 – 8/83: NSERC Summer Student

- *In vitro* metabolism of β -naphthoflavone in C57BL/6J and DBA mice

EDUCATION

Ph.D.

"Phenotypic and genotypic effects of mutations in the human retinoblastoma gene."

Supervisor: Dr. R. A. Phillips

University of Toronto
Toronto, Ontario, Canada.
Department of Medical
Biophysics.

1989

Honours B.Sc

"The *in vitro* metabolism of the cytochrome P-448 inducer β -naphthoflavone in C57BL/6J mice."

Supervisor: Dr. G. D. Sweeney

McMaster University,
Hamilton, Ontario, Canada.
Department of Biochemistry

1983

ACADEMIC AWARDS

| | |
|---|-----------|
| Imperial Cancer Research Fund Postdoctoral Fellowship | 1989-1992 |
| Medical Research Council Studentship | 1983-1988 |
| NSERC Undergraduate Summer Research Award | 1983 |
| Society of Chemical Industry Merit Award (Hons. Biochem.) | 1983 |
| Dr. Harry Lyman Hooker Scholarship | 1981-1983 |
| J.L.W. Gill Scholarship | 1981-1982 |
| Business and Professional Women's Club Scholarship | 1980-1981 |
| Wyerhauser Foundation Scholarship | 1979-1980 |

INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function, Litchfield Park, AZ, USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76th Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

Goddard A, Godowski PJ and Gurney AL. NL3 Tie ligand homologue nucleic acids. Patent Number: 6,426,218. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,4137,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date of Patent: May 14, 2002.

Goddard A, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homologue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 19, 2002.

Goddard A, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,873. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

PUBLICATIONS

Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD** Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.

Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausek SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.

Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.

Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.

Lee J, Ho WH, Maruoka M, Corpuz RT, Baldwin DT, Foster JS, **Goddard AD**, Yansura DG, Vandlen RL, Wood WI, Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.

Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.

Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8950-8954.

Guo S, Yamaguchi Y, Schilbach S, Wada T.; Lee J, **Goddard A**, French D, Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.

Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.

Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.

Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.

Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused; a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.

Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

- Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* **274**(15): 10287-10292.
- Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* **9**(4): 215-218.
- Ridgway JBB, Ng E, Kern JA, Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* **59**: 2718-2723.
- Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**(6712): 699-703.
- Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* **95**(25): 14717-14722.
- Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**(6699): 284-288.
- Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* **16**(7): 677-681.
- Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* **8**(9): 525-528.
- Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothed mutations in sporadic basal-cell carcinoma. *Nature.* **391**(6662): 90-92.
- Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology.* **7**(12): 1003-1006.
- Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* **19**: 15-26.
- Sheridan JP, Marsters SA, Pitti RM, Gurney A., Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* **277** (5327): 818-821.

Goddard AD, Dowd P, Chernausk S, Geffner M, Gertner J, Hintz R, Hopwood N, Kaplan S, Plotnick L, Rogol A, Rosenfield R, Saenger P, Mauras N, Hershkopf R, Angulo M and Attie, K. (1997) Partial growth hormone insensitivity: The role of growth hormone receptor mutations in idiopathic short stature. *J. Pediatr.* **131**: S51-55.

Klein RD, Sherman D, Ho WH, Stone D, Bennett GL, Moffat B, Vandlen R, Simmons L, Gu Q, Hongo JA, Devaux B, Poulsen K, Armanini M, Nozaki C, Asai N, **Goddard A**, Phillips H, Henderson CE, Takahashi M and Rosenthal A. (1997) A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. **387**(6634): 717-21.

Stone DM, Hynes M, Armanini M, Swanson TA, Gu Q, Johnson RL, Scott MP, Pennica D, **Goddard A**, Phillips H, Noll M, Hooper JE, de Sauvage F and Rosenthal A. (1996) The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**(6605): 129-34.

Marsters SA, Sheridan JP, Donahue CJ, Pitti RM, Gray CL, **Goddard AD**, Bauer KD and Ashkenazi A. (1996) Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF-kappa β . *Current Biology* **6**(12): 1669-76.

Rothe M, Xiong J, Shu HB, Williamson K, **Goddard A** and Goeddel DV. (1996) I-TRAF is a novel TRAF-interacting protein that regulates TRAF-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **93**: 8241-8246.

Yang M, Luoh SM, **Goddard A**, Reilly D, Henzel W and Bass S. (1996) The bglX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic beta-glucosidase. *Microbiology* **142**: 1659-65.

Goddard AD and Black DM. (1996) Familial Cancer in Molecular Endocrinology of Cancer. Waxman, J. Ed. Cambridge University Press, Cambridge UK, pp.187-215.

Treanor JJS, Goodman L, de Sauvage F, Stone DM, Poulson KT, Beck CD, Gray C, Armanini MP, Pollocks RA, Hefti F, Phillips HS, **Goddard A**, Moore MW, Buj-Bello A, Davis AM, Asai N, Takahashi M, Vandlen R, Henderson CE and Rosenthal A. (1996) Characterization of a receptor for GDNF. *Nature* **382**: 80-83.

Klein RD, Gu Q, **Goddard A** and Rosenthal A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**: 7108-7113.

Winslow JW, Moran P, Valverde J, Shih A, Yuan JQ, Wong SC, Tsai SP, **Goddard A**, Henzel WJ, Hefti F and Caras I. (1995) Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**: 973-981.

Bennett BD, Zeigler FC, Gu Q, Fendly B, **Goddard AD**, Gillett N and Matthews W. (1995) Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Htk. *Proc. Natl. Acad. Sci. USA* **92**: 1866-1870.

Huang X, Yuang J, **Goddard A**, Foulis A, James RF, Lernmark A, Pujol-Borrell R, Rabinovitch A, Somoza N and Stewart TA. (1995) Interferon expression in the pancreases of patients with type I diabetes. *Diabetes* **44**: 658-664.

Goddard AD, Yuan JQ, Fairbairn L, Dexter M, Borrow J, Kozak C and Solomon E. (1995) Cloning of the murine homolog of the leukemia-associated PML gene. *Mammalian Genome* **6**: 732-737.

- Goddard AD**, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.
- Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominantly brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.
- Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.
- Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.
- Goddard AD** and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.
- Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Fioretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.
- Goddard AD**, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S-119S.
- Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.
- Foulkes W, **Goddard A.** and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.
- Goddard AD**, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.
- Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.
- Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the β -subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.
- Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a NotI linking library from human chromosome 17q. *Genomics* **10**: 477-480.
- Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* **46**: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* **62**: 394-408.

Goddard AD, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* **37**: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* **340**: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

Goddard AD, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* **8**: 2082-2088.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1986) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci. USA* **83**: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-myc expression. *Nature* **322**: 555-557.

Goddard AD, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* **152**: 31-38.

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger¹, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. ¹Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR¹ to clinical diagnostics are well known^{2,3}, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases⁴ made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization^{5,6}, gel electrophoresis with or without use of restriction digestion^{7,8}, HPLC⁹, or capillary electrophoresis¹⁰. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing¹¹.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.¹², developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.¹³, developed an assay in which the endogenous 5' exonuclease activity of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to dsDNA¹⁴⁻¹⁶. As outlined in Figure 1, a prototypic PCR

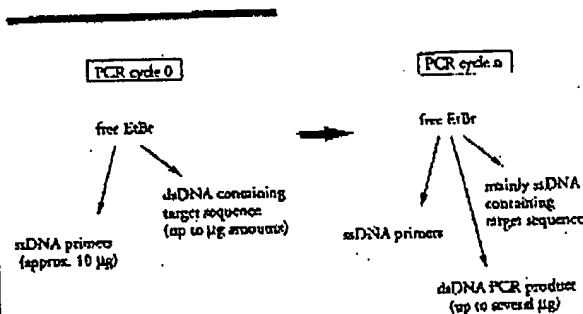


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

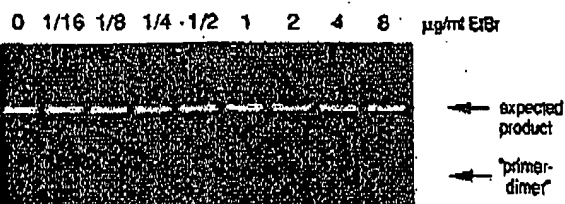


FIGURE 2 Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQα, made in the presence of increasing amounts of EtBr (up to 8 μg/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

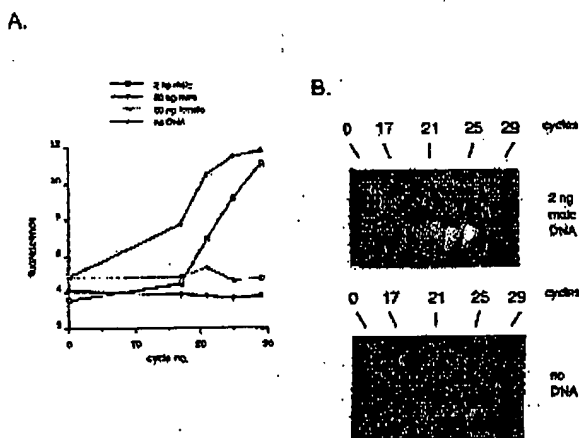


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μg/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photograph of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA¹⁷ to micrograms per PCR¹⁸. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the effect of EtBr in PCR, amplifications of the human HLA DQα gene¹⁹ were performed with the dye present at concentrations from 0.06 to 8.0 μg/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μg/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μg/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome²⁰. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β-globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μg/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β-globin gene²¹. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β-globin allele present^{21,22}.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β-globin individual (AA); from a heterozygous sickle β-globin individual (AS); and from a homozygous sickle β-globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

mocy.

ess the
HLA
cent at
oncen-
lowing
e 2, gel
ie yield
Br was
ndicat.

fic se-
nent of
ries of
rimers
human
either
DNA.
after 0,
or each
is fluo-
plotted
of this
case in
DNA is
umber.
cc-fold
ontain-
crease
her no
DNA
fewer
in fluo-
f these
the ex-
taining
in the

ualized
in a UV
b a red
ms that
A.
-globin
sch has
etection
Figure
ications
ography
These
for ci-
human
nparted
dual 3'
primer
has am-
c of the
ent^{21,22}
nsists of
the (left
Three
zygous,
ozygous
ozygous
genomic
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells⁶. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

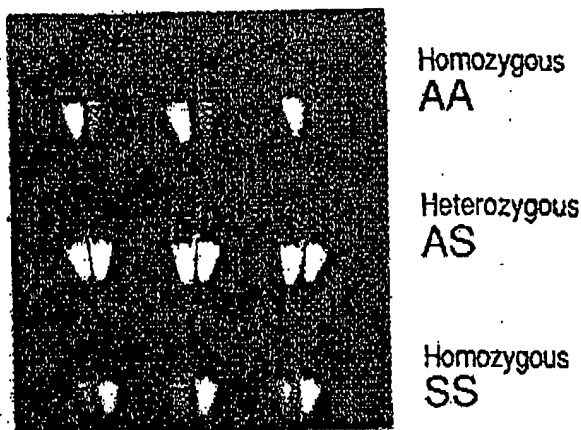


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

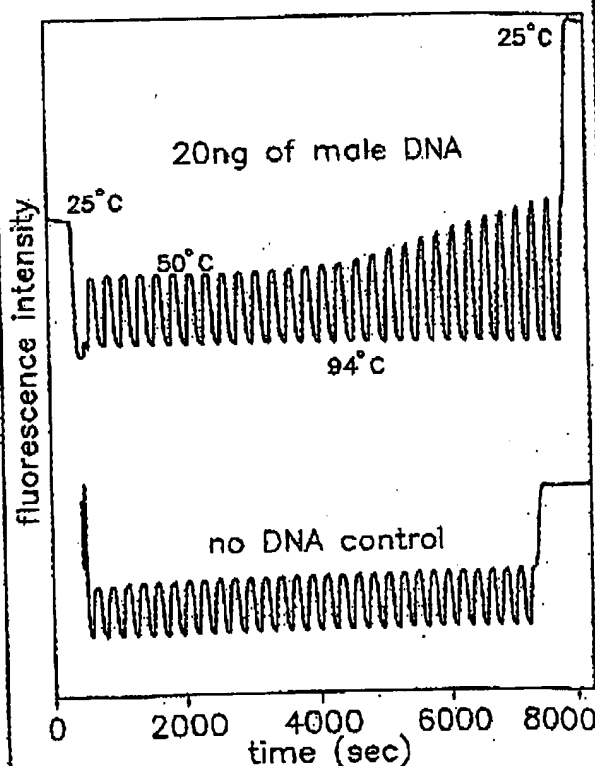


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube³, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins²³. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer²⁴.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format²⁵. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader²⁶.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiber optics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQ α gene amplifications containing EtBr. PCRs were set up in 100 μ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 2.5 units of Taq DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ α gene specific oligonucleotide primers GH26 and GH27¹⁹ and approximately 10^5 copies of DQ α PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

Y-chromosome specific PCR. PCRs (100 μ l total reaction volume) containing 0.5 μ g/ml EtBr were prepared as described for HLA-DQ α , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2²⁰, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β -globin gene PCR. Amplifications of 100 μ l volume using 0.5 μ g/ml of EtBr were prepared as described for HLA-DQ α above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.²¹. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.²¹ to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 29A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San-Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griest, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

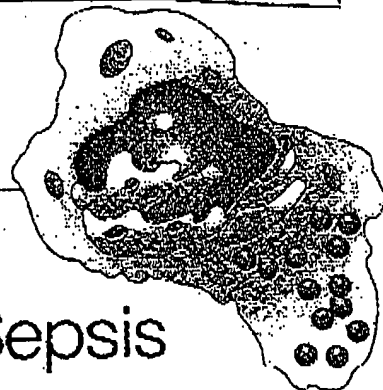
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQB* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Smitsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Fieroz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:293-294.
- Horn, G. T., Richards, B. and Klugger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Heiger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kim, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Feq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Szczer, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Seale, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ¹H NMR spectroscopy. *Nuc. Acids Res.* 18:3755-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:144-147.
- Abbott, M. A., Fieroz, B. J., Byrne, B. C., Kwok, S. Y., Smitsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. G., Doherty, M. and Gitocher, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Ugazoli, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spacie, D., Goda, L., Levenson, C. and Smitsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Submitted.*
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Picot, E., Williams, J. F. and Wendenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-105, 106-112.
- Tumosa, N. and Kahan, L. 1989. Fluorescent ELISA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

IBL

IMMUNO BIOLOGICAL LABORATORIES

sCD-14 ELISA

Trauma, Shock and Sepsis



The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations. During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips), precoated with a specific monoclonal antibody, 2x1 hour incubation, standard range: 3 - 96 ng/ml, detection limit: 1 ng/ml, CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH
OSTERSTRASSE 86 · D-2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98

BIOTECHNOLOGY VOL 10 APRIL 1992

417

GENENTECH, INC.
1 DNA Way
South San Francisco, CA 94080 USA
Phone: (650) 225-1000

FAX: (650) 952-9881

FACSIMILE TRANSMITTAL

Date: 19 July 2004

To: Anna Barry
Heller Ehrman

Re: Higuchi reference

Fax No: 324-6638

From: Patty Tobin, Assistant to Elizabeth M. Barnes, Ph.D.
Genentech, Inc. Legal Department

Number of Pages including this cover sheet: 6

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger¹, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. ¹Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR¹ to clinical diagnostics are well known^{2,3}, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases⁴ made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization^{5,6}, gel electrophoresis with or without use of restriction digestion^{7,8}, HPLC⁹, or capillary electrophoresis¹⁰. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing¹¹.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.¹², developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.¹³, developed an assay in which the endogenous 5' exonuclease activity of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to ds-DNA¹⁴⁻¹⁶. As outlined in Figure 1, a prototypic PCR

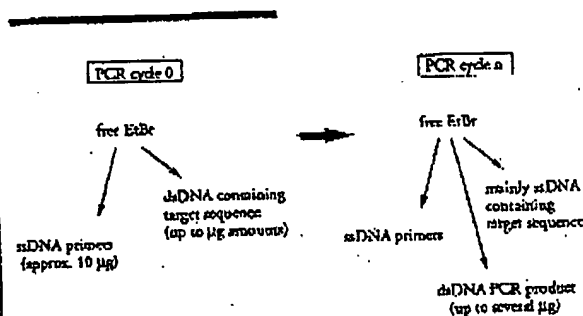


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

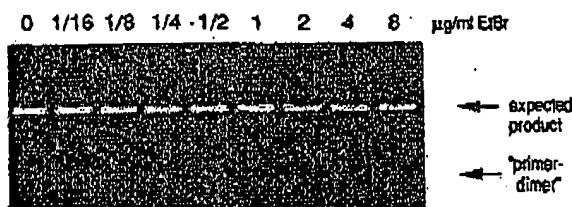
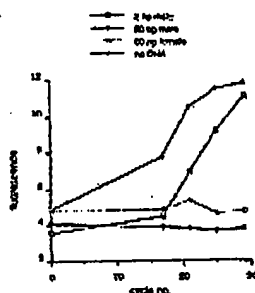


FIGURE 2 Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQA, made in the presence of increasing amounts of EtBr (up to 8 µg/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

A.



B.

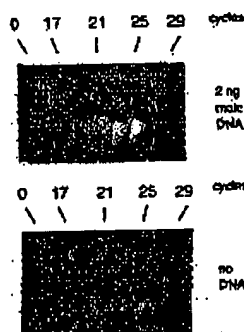


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 µg/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photograph of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA¹⁷ to micrograms per PCR¹⁸. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the effect of EtBr in PCR, amplifications of the human HLA DQA gene¹⁹ were performed with the dye present at concentrations from 0.06 to 8.0 µg/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 µg/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 µg/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome²⁰. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β -globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 µg/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β -globin gene²¹. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β -globin allele present^{21,22}.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β -globin individual (AA); from a heterozygous sickle β -globin individual (AS); and from a homozygous sickle β -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

emocy.

ess the
HLA
cent at
oncen-
lowing
e 2, gel
ic yield
Br was
indicat.

Se se-
nent of
ries of
rimers
human
either
DNA.
After 0,
or each
ts fluo-
plotted
of this
case in
DNA is
umber.
ce-fold
ontain-
ncrease
her no
DNA
fewer
in fluo-
f these
the ex-
taining
in the

ualized
n a UV
h a red
ns that
VA.

-globin
sch has
etection

Figure
ications
graphy
These
for ci-
human
nparted
nial 3'
primer
has am-
c of the
ent^{21,22}
nsists of
the (left
Three
zygous,
ozygous
ozygous
genomic
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells⁶. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

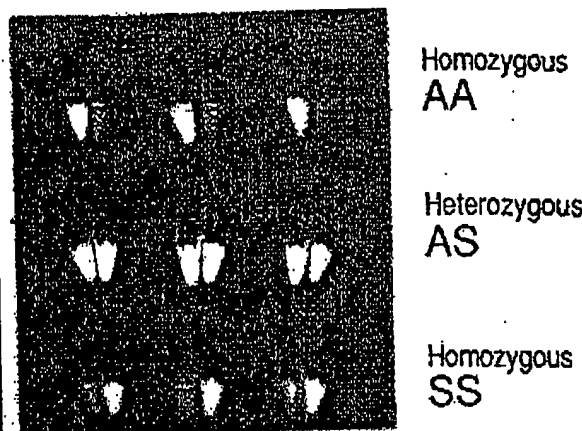


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCR's) for each input DNA.

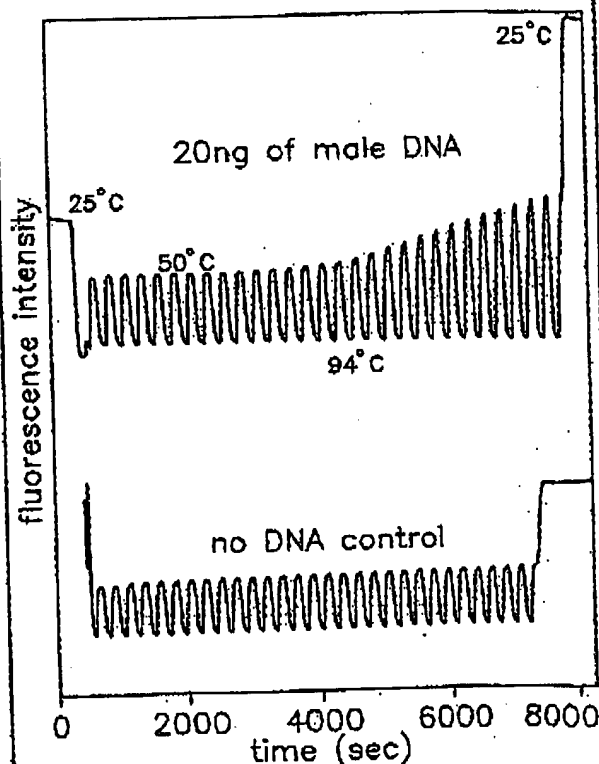


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube⁸, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins²³. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer²⁴.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format²⁵. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader²⁶.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQ α gene amplifications containing EtBr. PCRs were set up in 100 μ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ α gene specific oligonucleotide primers (CH26 and CH27¹⁹) and approximately 10^5 copies of DQ α PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

Y-chromosome specific PCR. PCRs (100 μ l total reaction volume) containing 0.5 μ g/ml EtBr were prepared as described for HLA-DQ α , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2²⁰, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β -globin gene PCR. Amplifications of 100 μ l volume using 0.5 μ g/ml of EtBr were prepared as described for HLA-DQ α above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.²¹. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.²¹ to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-96 transilluminator (UV-products San Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griest, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQB* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Sninsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligonucleotide cleavage detection. *J. Virol.* 61:1650-1654.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, Z. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:293-294.
- Horn, G. T., Richards, B. and Klugger, R. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Heiger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H.

1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Szczy, W. 1979. Interactions of 4',5'-diamidine-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ¹H NMR spectroscopy. *Nuc. Acids Res.* 18:3755-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:414-417.
- Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Sninsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1985. Analysis of enzymatically amplified β -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. C., Doherty, M. and Giuchieri, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Ugazochi, I., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Guda, L., Levenson, C. and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Submitted*.
- Higuchi, R. 1989. Using PCR to engineer DNA, p. 61-70. *In: PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Haff, L., Atwood, J. G., DiCesare, J., Katz, E., Pirozza, E., Williams, J. F. and Wondolberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-109, 106-112.
- Tumosa, N. and Kahan, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

IBL

IMMUNO BIOLOGICAL LABORATORIES

sCD-14 ELISA

Trauma, Shock and Sepsis

The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations.

During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features:

- 12x8 determinations (microtiter strips),
- precoated with a specific monoclonal antibody,
- 2x1 hour incubation,
- standard range: 3 - 96 ng/ml
- detection limit: 1 ng/ml
- CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH
OSTERSTRASSE 86 · D - 2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98

BIOTECHNOLOGY VOL 10 APRIL 1992

417

Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5' nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5' → 3' nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3'-end nucleotide. In all cases, the reporter dye was attached to the 5' end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3'-end nucleotide exhibited a larger signal in the 5' nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3'-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridization probes.

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.⁽¹⁾ The assay exploits the 5' → 3' nucleolytic activity of *Taq* DNA polymerase^(2,3) and is diagramed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5' end; and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).^(4,5) During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5' → 3' nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5' nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.⁽⁶⁾ Because of this, it has been assumed that the quencher dye must be attached near the 5' end. Surprisingly, we have found that attaching a rhodamine dye at the 3' end of a probe still provides adequate quenching for the probe to perform in the 5' nuclease

PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5' end and a quencher dye on the 3' end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

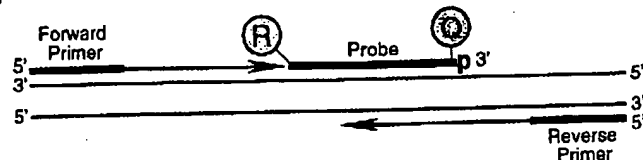
MATERIALS AND METHODS

Oligonucleotides

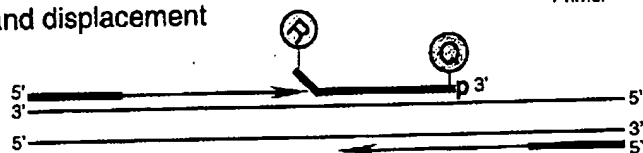
Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3'-blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5' end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3' end. Following deprotection and ethanol precipitation, TAMRA NHS ester was coupled to the LAN-containing oligonucleotide in 250

Research

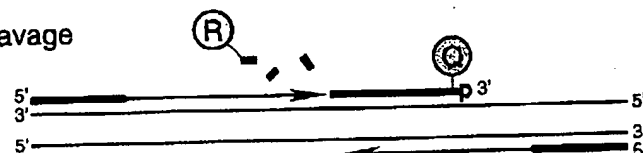
Polymerization



Strand displacement



Cleavage



Polymerization completed

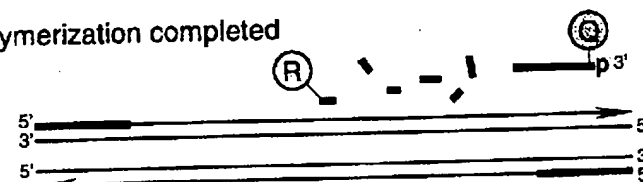


FIGURE 1 Diagram of 5' nuclease assay. Stepwise representation of the 5' → 3' nucleolytic activity of *Taq* DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mm Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C₈ 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the LAN-TAMRA moiety. For example, probe A1-7 has sequence A1 with LAN-TAMRA at nucleotide position 7 from the 5' end.

PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil N-glycosylase (Perkin-Elmer), and 1.25 unit of AmpliTaq DNA polymerase (Perkin-Elmer). A 295-bp segment from exon 3 of the human β-actin

gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.)⁽⁷⁾ was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.⁽⁸⁾ Actin amplification reactions contained 4 mM MgCl₂, 20 ng of human genomic DNA, 50 nM A1 or A3 probe, and 300 nM each

primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 3.5 mM MgCl₂, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer F119, and 200 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer TaqMan LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 485-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

| Name | Type | Sequence |
|------|------------|----------------------------------|
| F119 | primer | ACCCACAGGAAGTATGACCACTC |
| R119 | primer | ATGTCGGCTTCGGGCTGACGTTCTGC |
| P2 | probe | TCCGATTACTGATCGTIGCCAACCAAGTp |
| P2C | complement | GTAAGTGGTTGGCAACGATCAGTAATGCGATG |
| P5 | probe | CGGAATTGCTGGTATCTATGACAAGGATp |
| P5C | complement | TTCATCCTTGTATAGATACCAAGCAATCCG |
| AFP | primer | TCACCCACACTGTGCCATCTACGA |
| ARP | primer | CAGCGGAACCGCTCATTGCCAATGG |
| A1 | probe | ATGCCCTCCCATGCCCCTCTCGGp |
| A1C | complement | AGACGCAGGATGGCATGGGGGAGGGCATAC |
| A3 | probe | CGCCCTGGACTTCGAGCAAGAGATp |
| A3C | complement | CCATCTCTTGCTCGAAGTCCAGGGCGAC |

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5' → 3' direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used in the 5' nuclease assay, and complement used to hybridize to the corresponding probe. For the probes, the underlined base indicates a position where LAN with TAMRA attached was substituted for a T. (p) The presence of a 3' phosphate on each probe.

A1-2 RAQGGCCCTCCCCCATGCCATCCTGCGTp
 A1-7 RATGCCCQCCCCCATGCCATCCTGCGTp
 A1-14 RATGCCCTCCCCCAQGCCATCCTGCGTp
 A1-19 RATGCCCTCCCCCATGCCAQCCTGCGTp
 A1-22 RATGCCCTCCCCCATGCCATCCQCGTp
 A1-26 RATGCCCTCCCCCATGCCATCCTGCGQp

| Probe | 518 nm | | 582 nm | | RQ ⁻ | RQ ⁺ | Δ RQ |
|-------|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | no temp. | + temp. | no temp. | + temp. | | | |
| A1-2 | 25.5 \pm 2.1 | 32.7 \pm 1.9 | 38.2 \pm 3.0 | 38.2 \pm 2.0 | 0.67 \pm 0.01 | 0.86 \pm 0.06 | 0.19 \pm 0.06 |
| A1-7 | 53.5 \pm 4.3 | 395.1 \pm 21.4 | 108.5 \pm 6.3 | 110.3 \pm 5.3 | 0.49 \pm 0.03 | 3.58 \pm 0.17 | 3.09 \pm 0.18 |
| A1-14 | 127.0 \pm 4.9 | 403.5 \pm 19.1 | 109.7 \pm 5.3 | 93.1 \pm 6.3 | 1.16 \pm 0.02 | 4.34 \pm 0.15 | 3.18 \pm 0.15 |
| A1-19 | 187.5 \pm 17.9 | 422.7 \pm 7.7 | 70.3 \pm 7.4 | 73.0 \pm 2.8 | 2.67 \pm 0.05 | 5.80 \pm 0.15 | 3.13 \pm 0.16 |
| A1-22 | 224.6 \pm 9.4 | 482.2 \pm 43.6 | 100.0 \pm 4.0 | 96.2 \pm 9.6 | 2.25 \pm 0.03 | 5.02 \pm 0.11 | 2.77 \pm 0.12 |
| A1-26 | 160.2 \pm 8.9 | 454.1 \pm 18.4 | 93.1 \pm 5.4 | 90.7 \pm 3.2 | 1.72 \pm 0.02 | 5.01 \pm 0.08 | 3.29 \pm 0.08 |

FIGURE 2 Results of 5' nuclease assay comparing β -actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average \pm 1 s.d. for six reactions run without added template (no temp.) and six reactions run with template (+temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ⁻ and RQ⁺ values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, Δ RQ is calculated by subtracting the RQ value of the no-template control (RQ⁻) from the RQ value for the complete reaction including template (RQ⁺).

RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5' nuclease PCR assay. These probes hybridize to a target

sequence in the human β -actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β -actin gene containing the target sequence. Performance in the 5' nuclease PCR assay is monitored by the magnitude of Δ RQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a Δ RQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5' end, there is insufficient room for *Taq* polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable Δ RQ values that are

clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ⁻ depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ⁻ values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3' end to quench 6-FAM on the 5' end, three additional pairs of probes were tested in the 5' nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3' end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3' quencher exhibits a Δ RQ value that is considerably higher than for the probe with the internal quencher. The RQ⁻ values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3' end of an oligonucleotide can quench efficiently the

TABLE 2 Results of 5' Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3'-terminal Nucleotide

| Probe | 518 nm | | 582 nm | | RQ ⁻ | RQ ⁺ | Δ RQ |
|-------|-----------------|------------------|-----------------|------------------|-----------------|-----------------|-----------------|
| | no temp. | + temp. | no temp. | + temp. | | | |
| A3-6 | 54.6 \pm 3.2 | 84.8 \pm 3.7 | 116.2 \pm 6.4 | 115.6 \pm 2.5 | 0.47 \pm 0.02 | 0.73 \pm 0.03 | 0.26 \pm 0.04 |
| A3-24 | 72.1 \pm 2.9 | 236.5 \pm 11.1 | 84.2 \pm 4.0 | 90.2 \pm 3.8 | 0.86 \pm 0.02 | 2.62 \pm 0.05 | 1.76 \pm 0.05 |
| P2-7 | 82.8 \pm 4.4 | 384.0 \pm 34.1 | 105.1 \pm 6.4 | 120.4 \pm 10.2 | 0.79 \pm 0.02 | 3.19 \pm 0.16 | 2.40 \pm 0.16 |
| P2-27 | 113.4 \pm 6.6 | 555.4 \pm 14.1 | 140.7 \pm 8.5 | 118.7 \pm 4.8 | 0.81 \pm 0.01 | 4.68 \pm 0.10 | 3.88 \pm 0.10 |
| P5-10 | 77.5 \pm 6.5 | 244.4 \pm 15.9 | 86.7 \pm 4.3 | 95.8 \pm 6.7 | 0.89 \pm 0.05 | 2.55 \pm 0.06 | 1.66 \pm 0.08 |
| P5-28 | 64.0 \pm 5.2 | 333.6 \pm 12.1 | 100.6 \pm 6.1 | 94.7 \pm 6.3 | 0.63 \pm 0.02 | 3.53 \pm 0.12 | 2.89 \pm 0.13 |

Reactions containing the indicated probes and calculations were performed as described in Material and Methods and in the legend to Fig. 2.

Research

fluorescence of a reporter dye on the 5' end. The degree of quenching is sufficient for this type of oligonucleotide to be used as a probe in the 5' nuclease PCR assay.

To test the hypothesis that quenching by a 3' TAMRA depends on the flexibility of the oligonucleotide, fluorescence was measured for probes in the single-stranded and double-stranded states. Table 3 reports the fluorescence observed at 518 and 582 nm. The relative degree of quenching is assessed by calculating the RQ ratio. For probes with TAMRA 6–10 nucleotides from the 5' end, there is little difference in the RQ values when comparing single-stranded with double-stranded oligonucleotides. The results for probes with TAMRA at the 3' end are much different. For these probes, hybridization to a complementary strand causes a dramatic increase in RQ. We propose that this loss of quenching is caused by the rigid structure of double-stranded DNA, which prevents the 5' and 3' ends from being in proximity.

When TAMRA is placed toward the 3' end, there is a marked Mg^{2+} effect on quenching. Figure 3 shows a plot of observed RQ values for the A1 series of probes as a function of Mg^{2+} concentration. With TAMRA attached near the 5' end (probe A1-2 or A1-7), the RQ value at 0 mM Mg^{2+} is only slightly higher than RQ at 10 mM Mg^{2+} . For probes A1-19, A1-22, and A1-26, the RQ values at 0 mM Mg^{2+} are very high, indicating a much

reduced quenching efficiency. For each of these probes, there is a marked decrease in RQ at 1 mM Mg^{2+} followed by a gradual decline as the Mg^{2+} concentration increases to 10 mM. Probe A1-14 shows an intermediate RQ value at 0 mM Mg^{2+} with a gradual decline at higher Mg^{2+} concentrations. In a low-salt environment with no Mg^{2+} present, a single-stranded oligonucleotide would be expected to adopt an extended conformation because of electrostatic repulsion. The binding of Mg^{2+} ions acts to shield the negative charge of the phosphate backbone so that the oligonucleotide can adopt conformations where the 3' end is close to the 5' end. Therefore, the observed Mg^{2+} effects support the notion that quenching of a 5' reporter dye by TAMRA at or near the 3' end depends on the flexibility of the oligonucleotide.

DISCUSSION

The striking finding of this study is that it seems the rhodamine dye TAMRA, placed at any position in an oligonucleotide, can quench the fluorescent emission of a fluorescein (6-FAM) placed at the 5' end. This implies that a single-stranded, double-labeled oligonucleotide must be able to adopt conformations where the TAMRA is close to the 5' end. It should be noted that the decay of 6-FAM in the excited state requires a certain amount of time. Therefore, what

matters for quenching is not the average distance between 6-FAM and TAMRA but, rather, how close TAMRA can get to 6-FAM during the lifetime of the 6-FAM excited state. As long as the decay time of the excited state is relatively long compared with the molecular motions of the oligonucleotide, quenching can occur. Thus, we propose that TAMRA at the 3' end, or any other position, can quench 6-FAM at the 5' end because TAMRA is in proximity to 6-FAM often enough to be able to accept energy transfer from an excited 6-FAM.

Details of the fluorescence measurements remain puzzling. For example, Table 3 shows that hybridization of probes A1-26, A3-24, and P5-28 to their complementary strands not only causes a large increase in 6-FAM fluorescence at 518 nm but also causes a modest increase in TAMRA fluorescence at 582 nm. If TAMRA is being excited by energy transfer from quenched 6-FAM, then loss of quenching attributable to hybridization should cause a decrease in the fluorescence emission of TAMRA. The fact that the fluorescence emission of TAMRA increases indicates that the situation is more complex. For example, we have anecdotal evidence that the bases of the oligonucleotide, especially G, quench the fluorescence of both 6-FAM and TAMRA to some degree. When double-stranded, base-pairing may reduce the ability of the bases to quench. The primary factor causing the quenching of 6-FAM in an intact probe is the TAMRA dye. Evidence for the importance of TAMRA is that 6-FAM fluorescence remains relatively unchanged when probes labeled only with 6-FAM are used in the 5' nuclease PCR assay (data not shown). Secondary effectors of fluorescence, both before and after cleavage of the probe, need to be explored further.

Regardless of the physical mechanism, the relative independence of position and quenching greatly simplifies the design of probes for the 5' nuclease PCR assay. There are three main factors that determine the performance of a double-labeled fluorescent probe in the 5' nuclease PCR assay. The first factor is the degree of quenching observed in the intact probe. This is characterized by the value of RQ^- , which is the ratio of reporter to quencher fluorescent emissions for a no template control PCR. Influences on the value of RQ^- include the particular reporter and quencher

TABLE 3 Comparison of Fluorescence Emissions of Single-stranded and Double-stranded Fluorogenic Probes

| Probe | 518 nm | | 582 nm | | RQ | |
|-------|--------|--------|--------|--------|------|------|
| | ss | ds | ss | ds | ss | ds |
| A1-7 | 27.75 | 68.53 | 61.08 | 138.18 | 0.45 | 0.50 |
| A1-26 | 43.31 | 509.38 | 53.50 | 93.86 | 0.81 | 5.43 |
| A3-6 | 16.75 | 62.88 | 39.33 | 165.57 | 0.43 | 0.38 |
| A3-24 | 30.05 | 578.64 | 67.72 | 140.25 | 0.45 | 3.21 |
| P2-7 | 35.02 | 70.13 | 54.63 | 121.09 | 0.64 | 0.58 |
| P2-27 | 39.89 | 320.47 | 65.10 | 61.13 | 0.61 | 5.25 |
| P5-10 | 27.34 | 144.85 | 61.95 | 165.54 | 0.44 | 0.87 |
| P5-28 | 33.65 | 462.29 | 72.39 | 104.61 | 0.46 | 4.43 |

(ss) Single-stranded. The fluorescence emissions at 518 or 582 nm for solutions containing a final concentration of 50 nM indicated probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 10 mM $MgCl_2$. (ds) Double-stranded. The solutions contained, in addition, 100 nM A1C for probes A1-7 and A1-26, 100 nM A3C for probes A3-6 and A3-24, 100 nM P2C for probes P2-7 and P2-27, or 100 nM P5C for probes P5-10 and P5-28. Before the addition of $MgCl_2$, 120 μ l of each sample was heated at 95°C for 5 min. Following the addition of 80 μ l of 25 mM $MgCl_2$, each sample was allowed to cool to room temperature and the fluorescence emissions were measured. Reported values are the average of three determinations.

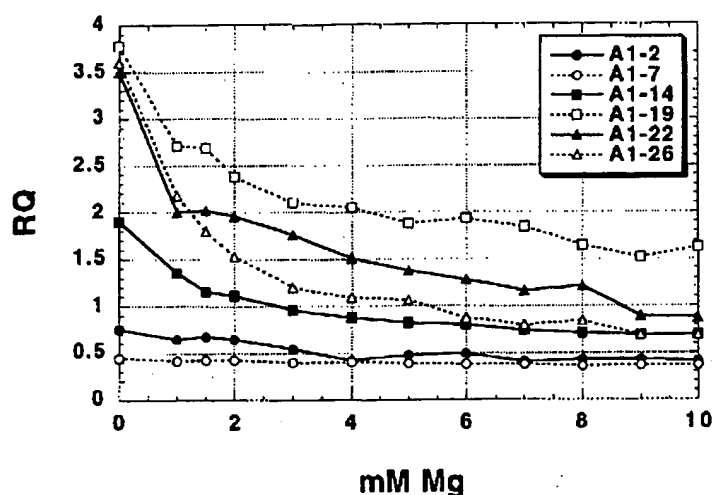


FIGURE 3 Effect of Mg^{2+} concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nm probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of $MgCl_2$. The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs. $MgCl_2$ concentration (mM Mg). The key (upper right) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe T_m , presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.⁽¹⁾

The rise in RQ^- values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed quencher. For the other three sets of

probes, the interpretation of RQ^- values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger RQ^- than the internal TAMRA probe. For the P2 pair, both probes have about the same RQ^- value. For the P5 probes, the RQ^- for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the RQ^- value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on RQ^- .

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes, ΔRQ is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as ΔRQ does not increase when the quencher is placed closer to the 3' end. This illus-

trates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the T_m of a probe. In fact, a 2°C–3°C reduction in T_m has been observed for two probes with internally attached TAMRAs.⁽⁹⁾ This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.⁽¹¹⁾ demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the $\Delta F508$ mutant. Their probes had TAMRA attached to the seventh nucleotide from

Research

the 5' end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3' TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.⁽¹⁰⁾ describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that requires adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* **21**: 3761-3766.
2. Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction prod-

- uct by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* **88**: 7276-7280.
3. Lyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260**: 778-783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* **2**: 55-75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*, pp. 303-339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* **58**: 719-726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Inter-species homology of sequences in the introns. *Proc. Natl. Acad. Sci.* **82**: 6133-6137.
8. du Breuil, R.M., J.M. Patel, and B.V. Mendelow. 1993. Quantitation of β -actin-specific mRNA transcripts using xeno-competitive PCR. *PCR Methods Applic.* **3**: 57-59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, R.L. Christensen, and E.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences: Poly-dA and poly-A detection. *Nucleic Acids Res.* **22**: 2424-2425.

Received December 20, 1994; accepted in revised form March 6, 1995.

THIS MATERIAL MAY BE PROTECTED
BY COPYRIGHT LAW (17 U.S. CODE)

GENOME METHODS

Real Time Quantitative PCR

Christian A. Heid,¹ Junko Stevens,² Kenneth J. Livak,² and
P. Mickey Williams^{1,3}

¹BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080; ²Genentech, Inc., 460 Point San Bruno Blvd., Foster City, California 94404

²Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Tan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genome quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Platack et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)-PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative

that it be used properly for quantitation (Rauymaekers 1995). Many early reports of quantitative PCR and RT-PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Perre 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT-PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as β -actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)-PCR, has been developed and is used widely for PCR quantitation. QC-PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Ptacek et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

³Corresponding author.

REAL TIME QUANTITATIVE PCR

RESULTS

PCR Product Detection in Real Time

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QC-PCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT-PCR analysis: (1) agarose gels, (2) fluorescent labeling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fusco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5' nuclease assay first described by Holland et al. (1991). The method uses the 5' nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Livak et al. 1995a,b). One fluorescent dye serves as a reporter [FAM (i.e., 6-carboxyfluorescein)] and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pF8TM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact, fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5'-3' nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human β -actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed ΔR_n (or ΔRQ) using the following equation: $\Delta R_n = (R_n') / (R_n'')$, where R_n' = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and R_n'' = emission intensity of re-

HUI ET AL.

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points (ΔRn s) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The ΔRn mean value is plotted on the y-axis, and time, represented by cycle number, is plotted on the x-axis. During the early cycles of the PCR amplification, the ΔRn

value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which

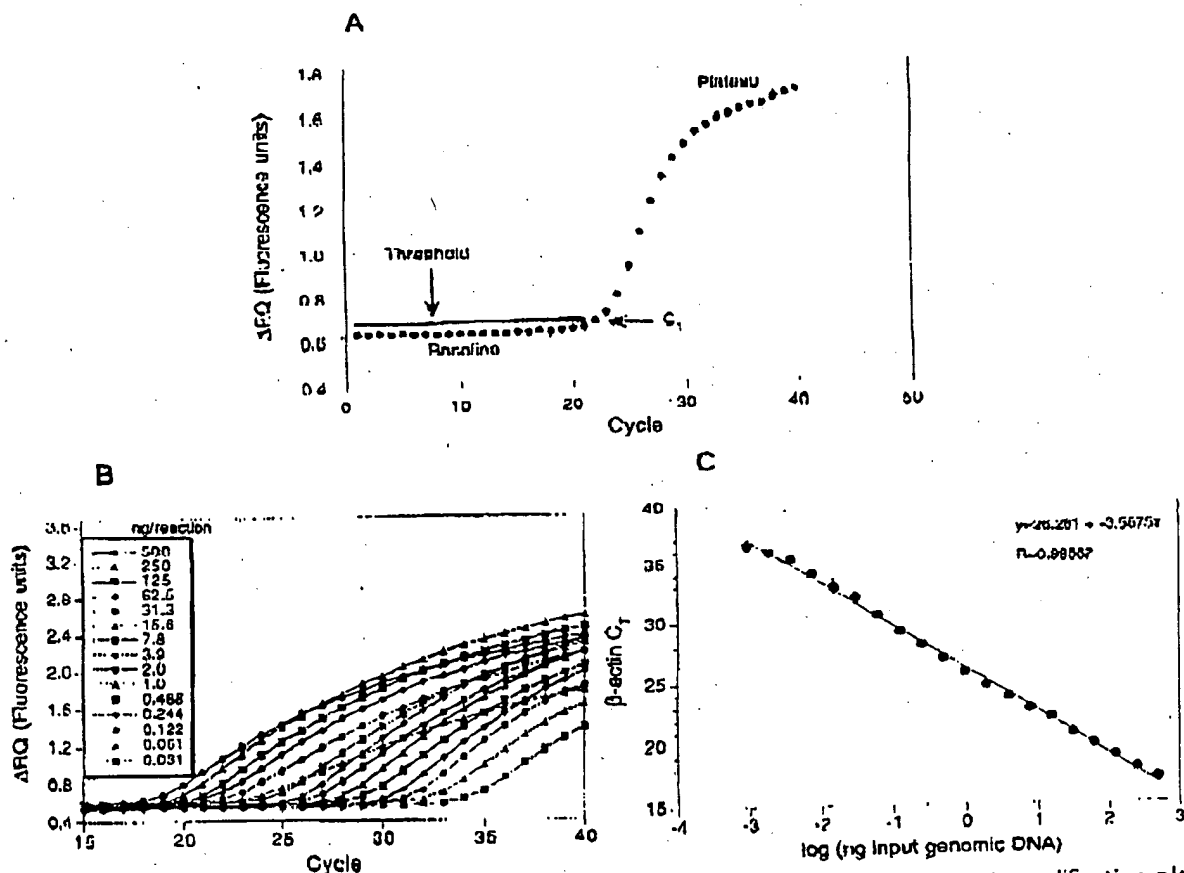


Figure 1 PCR product detection in real time. (A) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot. C_t values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (B) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with β -actin primers. (C) Input DNA concentration of the samples plotted versus C_t . All

REAL TIME QUANTITATIVE PCR

the amplification plot crosses the threshold is defined as C_T . C_T is reported as the cycle number at this point. As will be demonstrated, the C_T value is predictive of the quantity of input target.

C_T Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human β actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the C_T values. Figure 1C represents the C_T values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The C_T values decrease linearly with increasing target quantity. Thus, C_T values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated C_T value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar C_T values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using C_T values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Se-

ments over a very large range of relative starting target quantities.

Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing β -actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of C_T values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean C_T values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for β -actin gene quantity. The highest C_T difference between any of the samples was 0.85 and 0.71 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured β -actin C_T value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected C_T value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

Quantitative Analysis of a Plasmid After

PCR 001 RFR VVJ AC:BT 2002/00/27

FINAL

Table 1. Reproducibility of Sample Preparation Method

| Sample no. | 100 ng | | | | 25 ng | | | |
|------------|----------------|-------|--------------------|------|----------------|-------|--------------------|------|
| | C _T | mean | standard deviation | CV | C _T | mean | standard deviation | CV |
| 1 | 18.24 | 18.27 | 0.06 | 0.32 | 20.48 | 20.51 | 0.03 | 0.17 |
| | 18.23 | | | | 20.55 | | | |
| | 18.33 | | | | 20.5 | | | |
| 2 | 18.33 | 18.37 | 0.06 | 0.32 | 20.61 | 20.54 | 0.11 | 0.54 |
| | 18.35 | | | | 20.59 | | | |
| | 18.44 | | | | 20.41 | | | |
| 3 | 18.3 | 18.34 | 0.07 | 0.36 | 20.54 | 20.54 | 0.06 | 0.28 |
| | 18.3 | | | | 20.6 | | | |
| | 18.42 | | | | 20.49 | | | |
| 4 | 18.15 | 18.23 | 0.08 | 0.46 | 20.48 | 20.43 | 0.05 | 0.26 |
| | 18.23 | | | | 20.44 | | | |
| | 18.32 | | | | 20.38 | | | |
| 5 | 18.4 | 18.42 | 0.04 | 0.23 | 20.68 | 20.73 | 0.13 | 0.61 |
| | 18.38 | | | | 20.87 | | | |
| | 18.46 | | | | 20.63 | | | |
| 6 | 18.54 | 18.74 | 0.24 | 1.26 | 21.09 | 21.06 | 0.03 | 0.15 |
| | 18.67 | | | | 21.04 | | | |
| | 19 | | | | 21.01 | | | |
| 7 | 18.28 | 18.39 | 0.12 | 0.66 | 20.67 | 20.68 | 0.04 | 0.2 |
| | 18.36 | | | | 20.73 | | | |
| | 18.52 | | | | 20.65 | | | |
| 8 | 18.45 | 18.63 | 0.16 | 0.83 | 20.98 | 20.86 | 0.12 | 0.57 |
| | 18.7 | | | | 20.84 | | | |
| | 18.73 | | | | 20.75 | | | |
| 9 | 18.18 | 18.29 | 0.1 | 0.55 | 20.46 | 20.51 | 0.07 | 0.32 |
| | 18.34 | | | | 20.54 | | | |
| | 18.36 | | | | 20.48 | | | |
| 10 | 18.42 | 18.55 | 0.12 | 0.65 | 20.79 | 20.73 | 0.1 | 0.16 |
| | 18.57 | | | | 20.78 | | | |
| | 18.66 | | | | 20.62 | | | |
| Mean | (1 10) | 18.42 | 0.17 | 0.90 | | 20.66 | 0.19 | 0.94 |

for containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 µg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β-Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β-actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β-actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β-actin C_T values of the triplicates were plotted (error bars represent standard deviation). The highest difference

between any two sample means was 0.95 C_T. Ten nanograms of total DNA of each sample were also examined for β-actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β-actin C_T value difference was 1.0. As figure 3 shows, the rate of β-actin C_T change between the 100 and 10-ng samples was similar (slope values range between 3.56 and -3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished

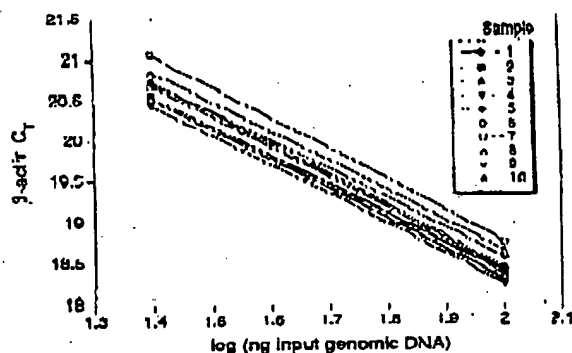


Figure 2 Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs. C_t . In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean β -actin C_t value obtained for each 100-ng sample on a β -actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample, a , was obtained by extrapolation to the x-axis.

Figure 4A shows the measured (i.e., non-normalized) quantities of factor VIII plasmid DNA (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate

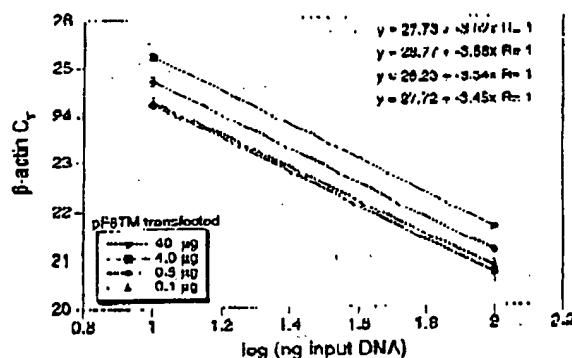


Figure 3 Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1 μ g of pF8TM) were analyzed for the β -actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the β -actin C_t values are plotted versus the total input DNA concentration.

REAL TIME QUANTITATIVE PCR

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean C_t values increase) with decreasing amounts of plasmid (transfected). The mean C_t values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM, b , found in each of the four transfections was determined by extrapolation to the x-axis of the standard curve in Figure 4B. These uncorrected values, b , for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100 \text{ ng}}{a} = \text{actual pF8TM copies per 100 ng of genomic DNA}$$

where a = actual genomic DNA in a sample and b = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40 μ g of plasmid, was 35 μ g per 100 ng genomic DNA. This results in ~520 plasmid copies per cell.

DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT-PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e., β -actin) or a "housekeeping" gene for RT-PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

HLID ET AL.

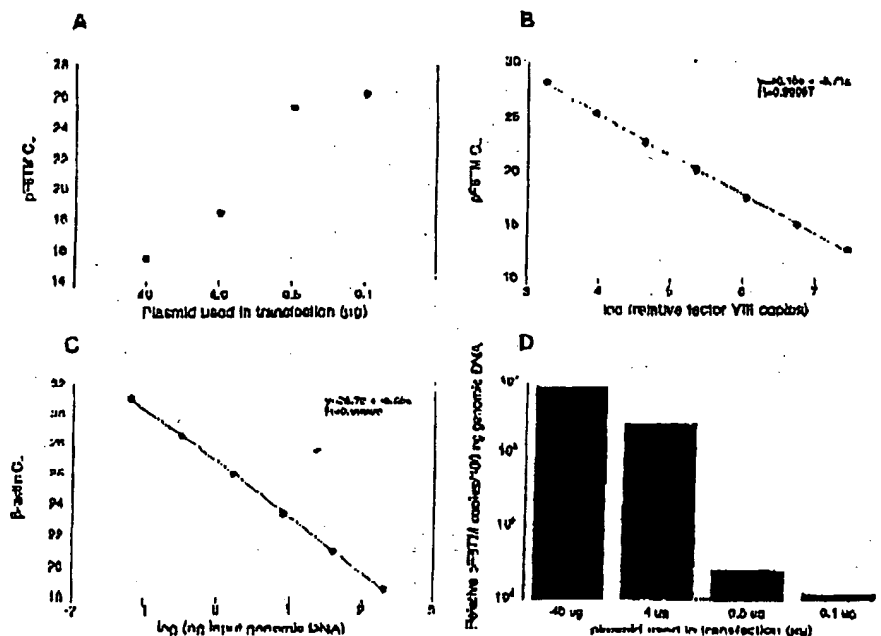


Figure 4 Quantitative analysis of pF8TM in transfected cells. (A) Amount of plasmid DNA used for the transfection plotted against the mean C_T value determined for pF8TM remaining 24 hr after transfection. (B,C) Standard curves of pF8TM and β-actin, respectively. pF8TM DNA (B) and genomic DNA (C) were diluted serially 1:5 before amplification with the appropriate primers. The β-actin standard curve was used to normalize the results of A to 100 ng of genomic DNA. (D) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e., β-actin) for quantitative PCR or house-keeping genes for quantitative RT-PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multi-gene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed

for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values, C_T, correlate linearly with relative DNA copy numbers. Real time quantitative RT-PCR methodology (Gibson et al., this issue) has also been developed. Finally, real time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications [quantitative gene expression (RT-PCR), gene copy assays (Her2, IIIV, etc.), genotyping (knockout mouse analysis), and immunoprecipitation].

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes—greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

METHODS

Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol B from Tel Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8c251 (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT-PCR (CionaAmp EZ cDNA RNA PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)) using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with *Hind*III and *Hind*III restriction site sequences at the 5' end) and cloned into pSILM-32 (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5'-CCGCGTTCACAGAGTGAATGTC-3' and F8rev 5'-AAACCTTCAACCTTGGATCGTACG-3'. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 5' untranslated region of the parent pCIS2.8c251 plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human β -actin gene was amplified with the primers β -actin forward primer 5'-TCACCCACACTCTTGCCTTCTACCA-3' and β -actin reverse primer 5'-CAACCGAACCCTTCATTGCAATGG-3'. The reaction produced a 295-bp PCR product.

Amplification reactions (50 μ l) contained a DNA sample, 10X PCR Buffer II (5 μ l), 200 μ M dATP, dCTP, dGTP, and 400 μ M dTTP, 4 mM MgCl₂, 1.25 Units AmpliTaq DNA polymerase, 0.5 unit Amplitase uracil N-glycosylase (UNG), 60 pmole of each factor VIII primer, and 15 pmole of each β -actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5'(FAM)AGCTTCCTCCACCTGCTTCCTTCTGTGCCTT(TAMRA)p-3' and β -actin probe 5'(FAM)ATGCCCX(TAMRA)CCCCCATGCCATCp-3' where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801 0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector Instrument manual. Briefly, probe T_m should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with

REAL TIME QUANTITATIVE PCR

reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM'S F12 without GHT, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1 μ g were added to 1.5 ml of a solution containing 0.125 M CaCl₂ and 1X HEPES. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO₂ for 24 hr, washed with PBS, and resuspended in PBS. The remaining cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200 μ l of 20 mM Tris-HCl at pH 8.0.

ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bassler, H.A., S.J. Flood, K.J. Livak, J. Marmaro, R. Kohn, and C.A. Ball. 1995. Use of a fluorogenic probe in a PCR-based assay for the detection of *Listeria monocytogenes*. *App. Environ. Microbiol.* 61: 3724-3728.
- Becker-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Meth. Mol. Cell. Biol.* 2: 189-201.
- Clementi, M., S. Menzo, P. Bugnarelli, A. Manzi, A. Valenza, and P.E. Varaldo. 1997. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 197-196.
- Connor, R.J., H. Mohr, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4⁺ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772-1777.
- Eaton, D.L., W.J. Wood, D. Eaton, P.E. Hass, P.

HFID LI AL

Venar, and C. Gorman. 1986. Construction and characterization of an active factor VIII variant lacking the central one third of the molecule. *Biochemistry* 25: 8343-8347.

Basco, M.J., C.P. Treanor, S. Spivack, H.L. Wigge, and I.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140-147.

Ferre, J. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Applic.* 2: 1-9.

Bortado, M.R., L.A. Kingsley, and S.M. Wollinsky. 1995. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 69: 2092-2100.

Gibson, U.E.M., C.A. Heid, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).

Gorman, C.M., D.R. Gies, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3-10.

Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biochemistry* 10: 413-417.

Holland, P.M., R.D. Abramson, R. Watson, and D.J. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276-7280.

Huang, S.K., H.Q. Xiao, T.J. Klein, G. Paciotti, H.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688-2694.

Huang, S.K., M. Yi, E. Palmer, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen. *Annals N.Y. Acad. Sci.* 754: 6157-6162.

Kellogg, D.E., J.J. Srinisky, and S. Kowk. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202-208.

Lee, J.-G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761-3766.

Livak, K.J., S.J. Flood, J. Marmaro, W. Chuai, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic.* 4: 357-362.

Livak, K.J., J. Marmaro, and J.A. Todd. 1995b. Towards

fully automated genome-wide polymorphism screening [Letter]. *Nature Genet.* 9: 341-342.

Mulder, J., N. McKinney, C. Christopherson, J. Srinisky, L. Greenfield, and S. Kwok. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292-300.

Pang, S., Y. Koyanagi, S. Miller, C. Wiloy, H.V. Vinters, and I.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85-89.

Platak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1993a. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Aid Techniques* 14: 70-81.

Platak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hann, G.M. Shaw, and J.D. Lifson. 1993b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR [see Comments]. *Science* 259: 1749-1754.

Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495-503.

Racynackers, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 93-94.

Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750-768.

Slamon, D.J., G.M. Clark, S.G. Wong, W.J. Levin, A. Ullrich, and W.J. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177-182.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.

Tan, X., X. Sun, C.F. Gonzalez, and W. Hsueh. 1994. PAF and TNF increase the precursor of NF-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1215: 157-162.

Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201-5205.

Williams, S., C. Schwer, A. Krishnasao, C. Held, B. Karger, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (in press).

Received June 3, 1996; accepted in revised form July 29, 1996.

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABELL, ROBERT L. COHEN*, MONA F. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,††, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, †Molecular Biology, ‡Scientific Computing, and §Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 β (GSK-3 β) resulting in an increase in β -catenin levels. Stabilized β -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β -catenin levels (9). APC is phosphorylated by GSK-3 β , binds to β -catenin, and facilitates its degradation. Mutations in either APC or β -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0
PNAS is available online at www.pnas.org.

cDNA was synthesized from 2 μ g of poly(A)⁺ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μ g of poly(A)⁺ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

In Situ Hybridization. ³³P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNeasy.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{-\Delta Ct}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The δ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of $\approx 40,000$ (M_r 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of $\approx 27,000$ (M_r 27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

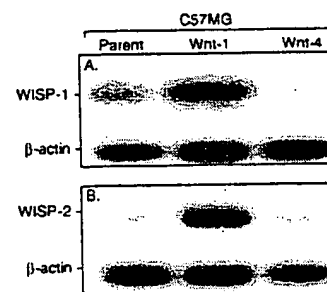


Fig. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)⁺ RNA (2 μ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human β -actin probe.

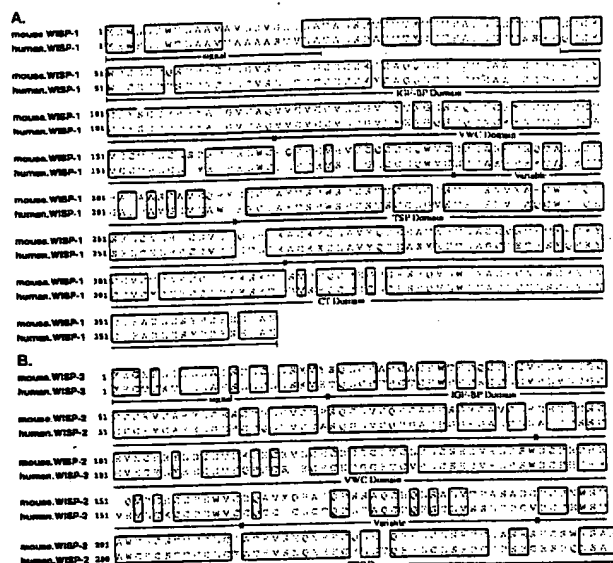


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

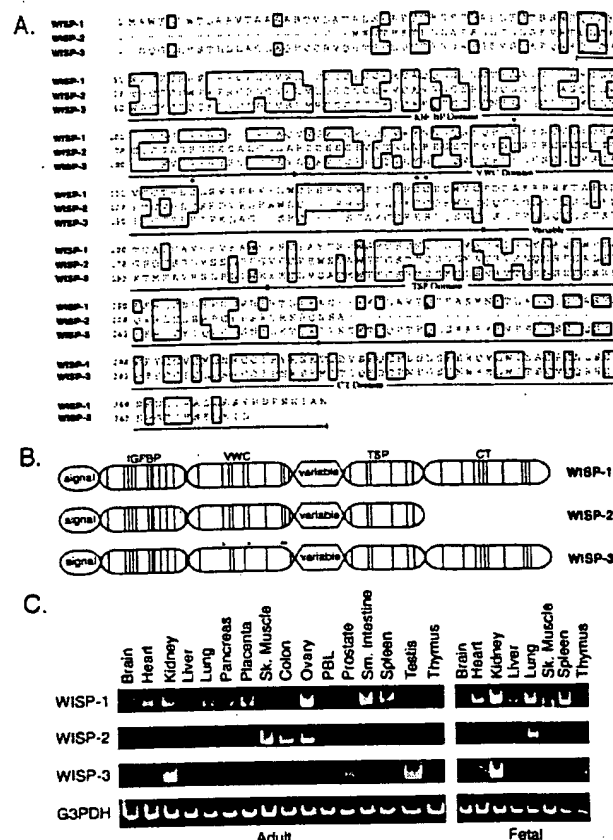


FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSXCSXCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

In Situ Localization of *WISP-1* and *WISP-2*. Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

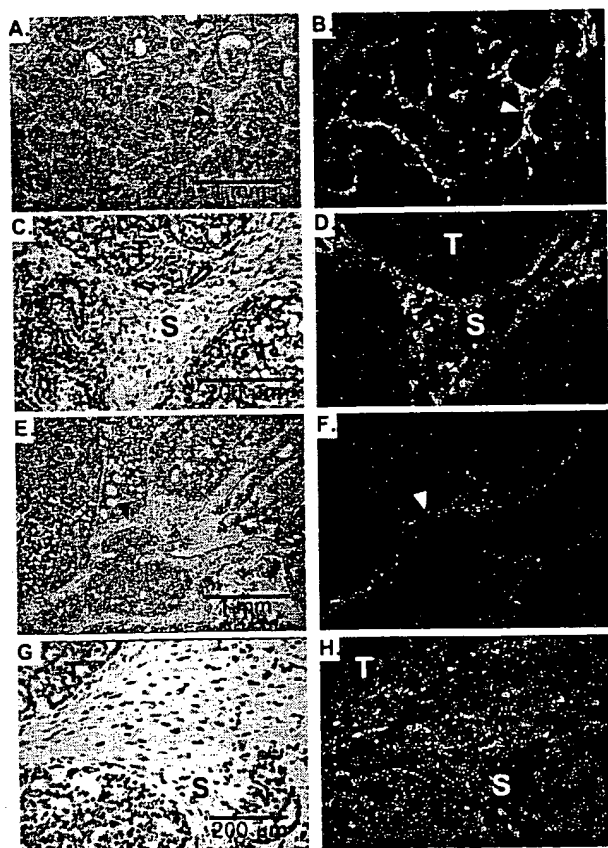


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

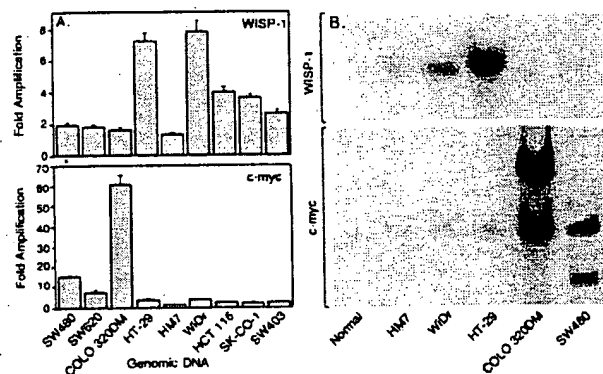


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10 μ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

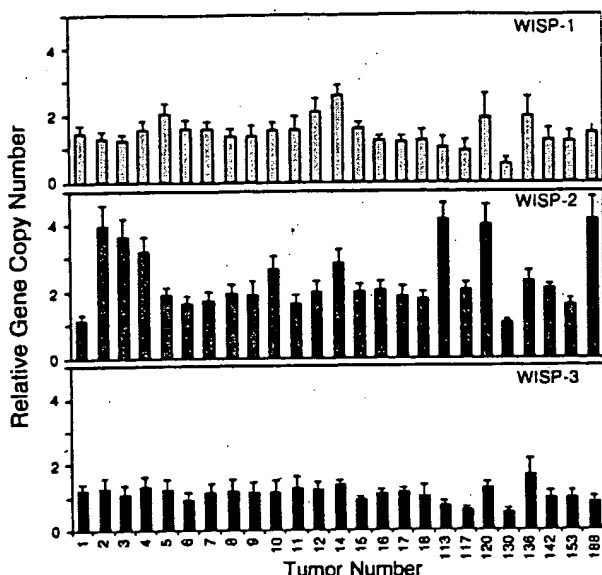


Fig. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

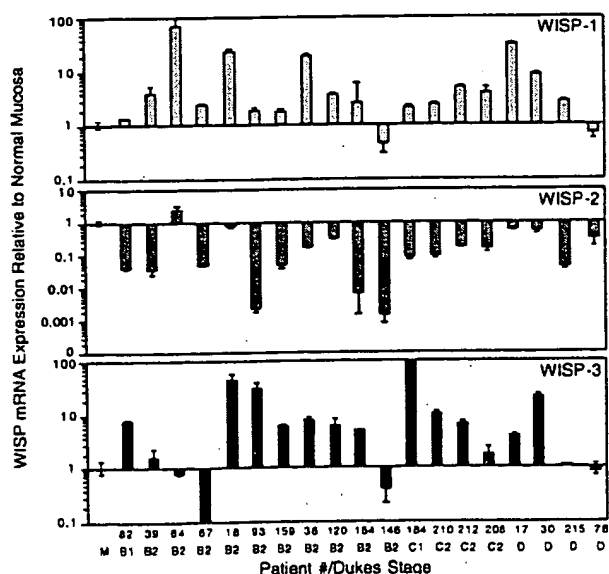


Fig. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- β , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- β 1, which is the stimulus for stromal proliferation (34). TGF- β 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISPs*. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISPs* as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISPs* in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
- Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
- Tsakamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182–187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248–1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278–6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131–6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326–1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
- Bork, P. (1993) *FEBS Lett.* 327, 125–130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10–21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195–205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell Genet.* 57, 109–111.
- Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
- Garte, S. J. (1993) *Crit. Rev. Oncol.* 4, 435–449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
- Wernert, N. (1997) *Virchows Arch.* 430, 433–443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., et al. (1994) *Cancer Res.* 54, 4257–4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., et al. (1998) *EMBO J.* 17, 3052–3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169–176.

methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described^{20,21}. Briefly, after antigen pulsing (30 µg ml⁻¹ TTCF) with tetrapeptides (1–2 mg ml⁻¹), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1 M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with 1 µCi of ³H-thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating 200 µg TTCF with 0.25 µg pig kidney legumain in 500 µl 50 mM citrate buffer, pH 5.5, for 1 h at 37 °C. **Glycopeptide digestions.** The peptides HIDNEEDI, HIDN(N-glucosamine) EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQLFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQLFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography¹¹. Glycopeptides corresponding to residues 622–642 and 421–452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at 30 °C with 5–50 mU ml⁻¹ pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of 10 mg ml⁻¹ α-cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

- Chen, J. M. et al. Cloning, isolation, and characterisation of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090–8098 (1997).
- Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208–213 (1993).
- Dalton, J. P., Hala Jarmiska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575–580 (1995).
- Bennett, K. et al. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1519–1524 (1992).
- Riese, R. J. et al. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357–366 (1996).
- Rodríguez, G. M. & Dimant, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894–2898 (1992).
- Hewitt, E. W. et al. Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693–4699 (1997).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821–850 (1997).
- Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93–102 (1998).
- Fineschi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377–382 (1997).
- Lu, J. & van Halbeek, H. Complete ¹H and ¹³C resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1–21 (1996).
- Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50–54 (1996).
- Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295–298 (1997).
- Wyatt, R. et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705–711 (1998).
- Boirelli, P. et al. N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128–3132 (1991).
- Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101–4109 (1990).
- Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535–559 (1981).
- Makoff, A. J., Ballantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043–1046 (1989).
- Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
- Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537–539 (1985).
- Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543–553 (1997).

Acknowledgements. We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti[†], Scot A. Marsters[†], David A. Lawrence[†], Margaret Roy^{*}, Frank C. Kischkel^{*}, Patrick Dowd^{*}, Arthur Huang^{*}, Christopher J. Donahue^{*}, Steven W. Sherwood^{*}, Daryl T. Baldwin^{*}, Paul J. Godowski^{*}, William I. Wood^{*}, Austin L. Gurney^{*}, Kenneth J. Hillan^{*}, Robert L. Cohen^{*}, Audrey D. Goddard^{*}, David Botstein[†] & Avi Ashkenazi^{*}

^{*} Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

[†] Department of Genetics, Stanford University, Stanford, California 94305, USA

[†] These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells¹. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (DcR3), that binds to FasL and inhibits FasL-induced apoptosis. The DcR3 gene was amplified in about half of 35 primary lung and colon tumours studied, and DcR3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily². Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (DcR3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)³, DcR3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of DcR3 in mammalian cells; DcR3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). DcR3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of DcR3 and OPG are conserved; however, the carboxy-terminal portion of DcR3 is 101 residues shorter.

We analysed expression of DcR3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high DcR3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of DcR3, we generated a recombinant, Fc-tagged DcR3 protein. We tested binding of DcR3-Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). DcR3-Fc showed a significant increase in binding to cells transfected with FasL⁴ (Fig. 2a), but not to cells transfected with TNF⁵, Apo2L/TRAIL^{6,7}, Apo3L/TWEAK^{8,9}, or OPGL/TRANCE/

RANKL¹⁰⁻¹² (data not shown). DcR3-Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3-Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3-Fc and Fas-Fc bound to soluble FasL with a comparable affinity ($K_d = 0.8 \pm 0.2$ and 1.1 ± 0.1 nM, respectively; Fig. 2e), and that DcR3-Fc could block nearly all of the binding of soluble FasL to Fas-Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3-Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3-Fc and Fas-Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at $\sim 0.1 \mu\text{g ml}^{-1}$. Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3-Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process¹. Consistent with previous results¹³, activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas-Fc blocked this effect substantially (Fig. 3c); DcR3-Fc blocked the

induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes¹⁴⁻¹⁶. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3-Fc and Fas-Fc each reduced killing of target cells from $\sim 65\%$ to $\sim 30\%$, with half-maximal inhibition at $\sim 1 \mu\text{g ml}^{-1}$; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3-Fc and Fas-Fc concentrations were required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL¹⁷.

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain

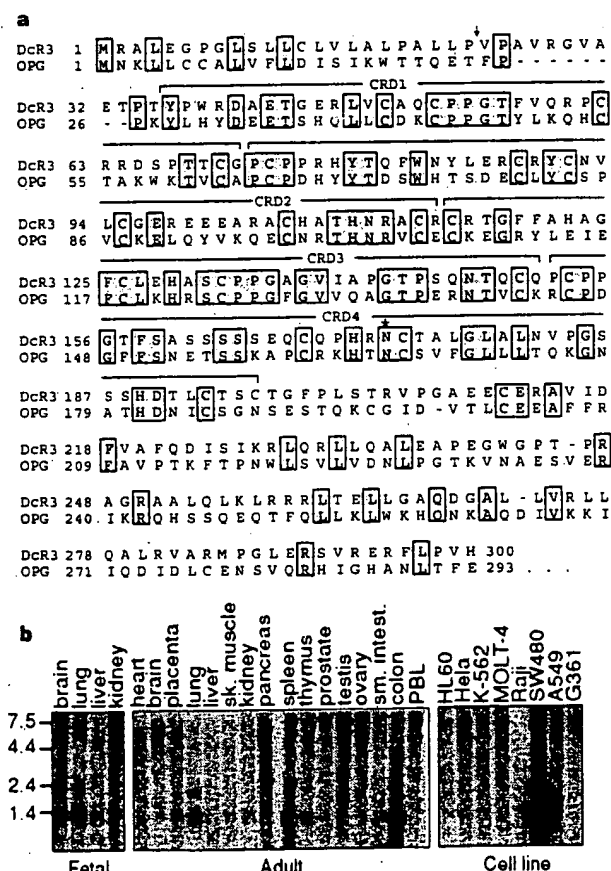


Figure 1 Primary structure and expression of human DcR3. **a**, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1-4), and the N-linked glycosylation site (asterisk) are shown. **b**, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)⁺ RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.

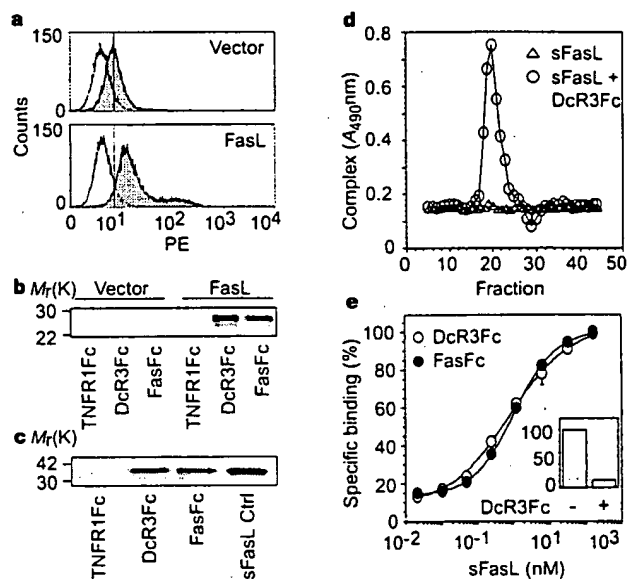


Figure 2 Interaction of DcR3 with FasL. **a**, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3-Fc (solid line, shaded area), TNFR1-Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ($P < 0.001$) between the binding of DcR3-Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. **b**, 293 cells were transfected as in **a** and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. **c**, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1-Fc, DcR3-Fc or Fas-Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. **d**, Flag-tagged sFasL was incubated with DcR3-Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3-Fc and sFasL-Flag. **e**, Equilibrium binding of DcR3-Fc or Fas-Fc to sFasL-Flag. Inset, competition of DcR3-Fc with Fas-Fc for binding to sFasL-Flag.

reaction (PCR)¹⁸ in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this,

we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epi-centre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG^{2,19}.

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas²⁰. A second mechanism involves proteolytic shedding of FasL from the cell surface¹⁷. DcR3 competes with Fas for

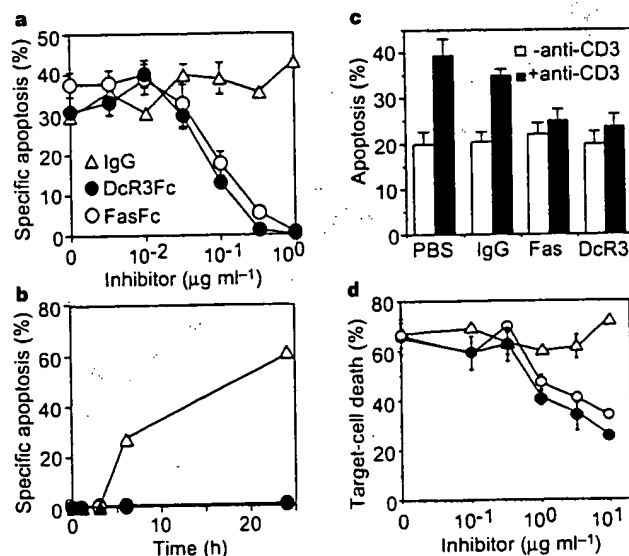


Figure 3 Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml⁻¹) oligomerized with anti-Flag antibody (0.1 μg ml⁻¹) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL-Flag plus anti-Flag antibody as in **a**, in presence of 1 μg ml⁻¹ DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml⁻¹). After 16 h, apoptosis of CD4⁺ cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with ⁵¹Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of ⁵¹Cr (mean ± s.d. for two donors, each in triplicate).

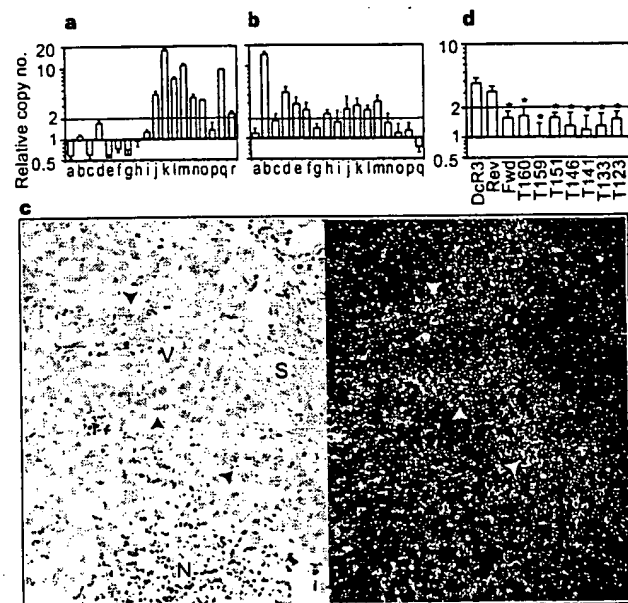


Figure 4 Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (b). Data are from two experiments done in duplicate. Asterisk indicates $P < 0.01$ for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described²¹. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L²². Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG³, which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L¹⁹. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response². Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

Methods

Isolation of DcR3 cDNA. Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in LifeseqTM (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

Fc-fusion proteins (immunoadhesins). The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described²³.

Fluorescence-activated cell sorting (FACS) analysis. We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL⁴ (2 µg), together with pRK5 encoding CrmA (2 µg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3-Fc or TNFR1-Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov-Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3-Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

Immunoprecipitation. Human 293 cells were transfected as above, and metabolically labelled with [³⁵S]cysteine and [³⁵S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 µM), the medium was immunoprecipitated with DcR3-Fc, Fas-Fc or TNFR1-Fc (5 µg), followed by protein A-Sepharose (Repligen). The precipitates were resolved by SDS-PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 µg) (Alexis) was incubated with each Fc-fusion protein (1 µg), precipitated with protein A-Sepharose, resolved by SDS-PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

Analysis of complex formation. Flag-tagged soluble FasL (25 µg) was incubated with buffer or with DcR3-Fc (40 µg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3-Fc-FasL complex in each fraction was analysed by placing 100 µl aliquots into microtitre wells pre-coated with anti-human IgG (Boehringer) to capture DcR3-Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin-horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3-Fc homodimers to two soluble FasL homotrimers.

Equilibrium binding analysis. Microtitre wells were coated with anti-human

IgG, blocked with 2% BSA in PBS. DcR3-Fc or Fas-Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas-Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3-Fc.

T-cell AICD. CD3⁺ lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotech), stimulated with phytohaemagglutinin (PHA; 2 µg ml⁻¹) for 24 h, and cultured in the presence of interleukin-2 (100 U ml⁻¹) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4⁺ cells²⁴.

Natural killer cell activity. Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotech), and incubated for 16 h with ⁵¹Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3-Fc, Fas-Fc or human IgG1. Target-cell death was determined by release of ⁵¹Cr in effector-target co-cultures relative to release of ⁵¹Cr by detergent lysis of equal numbers of Jurkat cells.

Gene-amplification analysis. Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR¹⁸ using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5'-CTTCTTCGCGCAGCTG-3' and 5'-ATCACGCCGCCACCA-3' and the fluorogenic probe sequence was 5'-(FAM-ACACGATGCGTGCTCCAGCAG AAp-(TAMARA), where FAM is 5'-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula 2^(ΔCT), where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

- Nagata, S. Apoptosis by death factor. *Cell* 88, 355-365 (1997).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959-962 (1994).
- Simonet, W. S. et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309-319 (1997).
- Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169-1178 (1993).
- Pennica, D. et al. Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724-729 (1984).
- Pitti, R. M. et al. Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687-12690 (1996).
- Wiley, S. R. et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673-682 (1995).
- Marsters, S. A. et al. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
- Chicheportiche, Y. et al. TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401-32410 (1997).
- Wong, B. R. et al. TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190-25194 (1997).
- Anderson, D. M. et al. A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175-179 (1997).
- Lacey, D. L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165-176 (1998).
- Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo1/(Fas/CD95). *Nature* 373, 438-441 (1995).
- Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235-1238 (1995).
- Medvedev, A. E. et al. Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394-404 (1997).
- Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13-18 (1997).
- Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31-36 (1998).
- Gelmini, S. et al. Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752-758 (1997).
- Emery, J. G. et al. Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363-14367 (1998).
- Wallach, D. Placing death under control. *Nature* 388, 123-125 (1997).
- Colotta, F. et al. Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* 261, 472-475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* 281, 1305–1308 (1998).
23. Ashkenazi, A. & Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* 9, 195–200 (1997).
24. Masters, S. et al. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

Crystal structure of the ATP-binding subunit of an ABC transporter

Li-Wei Hung*, Iris Xiaoyan Wang†, Kishiko Nikaido‡, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim*‡

* E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes¹. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters². Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains¹. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*^{3,4–8} is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, HisQMP₂, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins⁹, is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM⁶. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis⁵, the requirement for both subunits to be present for activity⁷, and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer³. HisP has been purified and characterized in an active soluble form³ which can be reconstituted into a fully active membrane-bound complex⁸.

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded β -sheet ($\beta 3$ and $\beta 8$ – $\beta 12$) spans both arms of the L, with a domain of a α - plus β -type structure ($\beta 1$, $\beta 2$, $\beta 4$ – $\beta 7$, $\alpha 1$ and $\alpha 2$) on one side (within arm I) and a domain of mostly α -helices ($\alpha 3$ – $\alpha 9$) on the

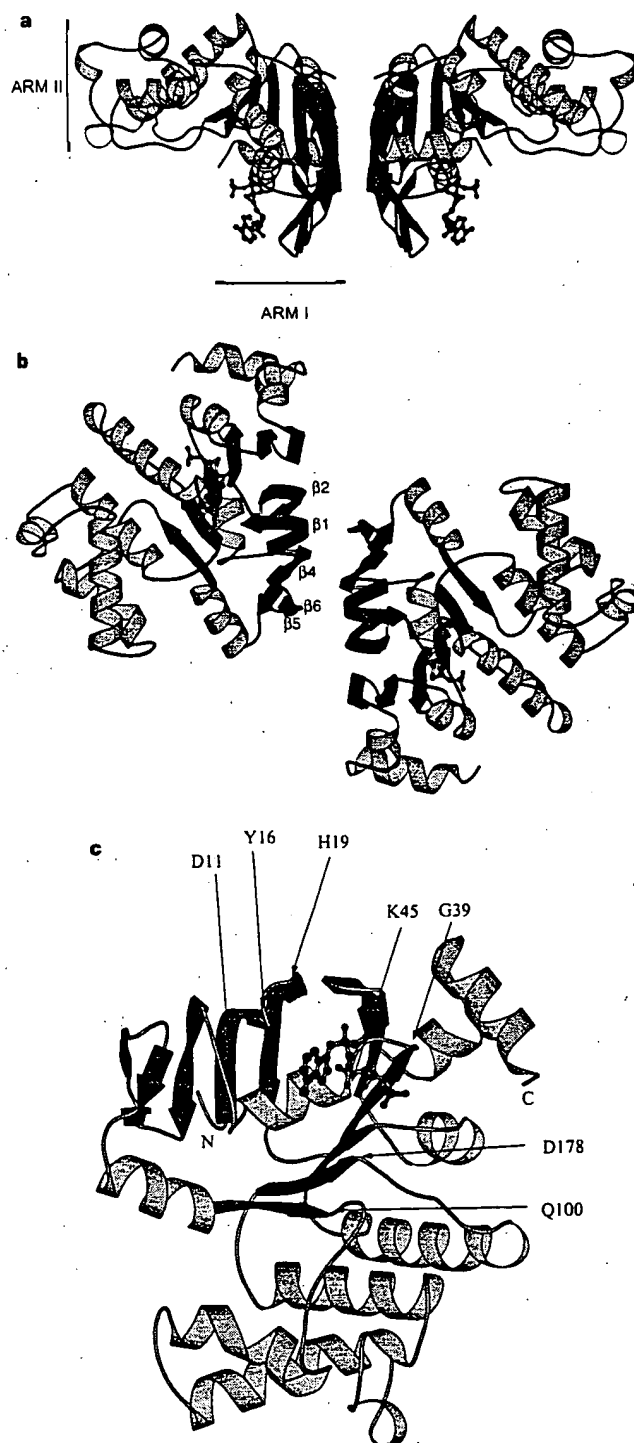


Figure 1 Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane. α -Helices are shown in orange and β -sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The β -strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a**–**c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT²⁸. N, amino terminus; C, C terminus.

NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE^{1,2}, Martine OLIV¹, Marie-Hélène CHAMPÈME², Dominique VIDAUD¹, Rosette LIDEREAU² and Michel VIDAUD^{1*}

¹Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France

²Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France.

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erbB2*) in breast tumors. Extra copies of *myc*, *ccnd1* and *erbB2* were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNAs; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.

© 1998 Wiley-Liss, Inc.

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmins) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erbB2* (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erbB2* proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuurin *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy

surrounding the link suggested by Slamon *et al.* (1987) between *erbB2* amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5' nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5' end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3' end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-4407-1754. E-mail: mvidaud@teaser.fr

Received 2 May 1998; Revised 30 June 1998

cycle, the fluorescent hybridization probe is hydrolyzed by the 5'-3' nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (i.e., ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the R_n (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): (i) the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridization of both the primers and the probe is necessary to generate a signal; (ii) the C_t (threshold cycle) value used for quantification is measured when PCR amplification is still in the log phase of PCR product accumulation. This is the main reason why C_t is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; (iii) use of C_t values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; (iv) The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; (v) the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; (vi) the assay is rapid, as results are available just one minute after thermal cycling is complete; (vii) the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erbB2*), as well as 2 genes (*alb* and *app*) located in a chromosome region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

MATERIAL AND METHODS

Tumor and blood samples

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguenin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

Real-time PCR

Theoretical basis. Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter C_t (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring C_t and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (i.e., lack

of extensive degradation) are both difficult to assess. We therefore also quantified a control gene (*alb*) mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast-tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (app, myc, ccnd1, erbB2)}}{\text{copy number of reference gene (alb)}}$$

Primers, probes, reference human genomic DNA and PCR consumables. Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAgency (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

Standard-curve construction. The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatk *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophoresed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/ μ l. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from 10^{-7} (10^3 copies of each gene) to 10^{-10} (10^0 copies). This series of diluted PCR products was aliquoted and stored at -80°C until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

PCR amplification. Amplification mixes (50 μ l) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes), $10\times$ TaqMan buffer (5 μ l), 200 μ M dATP, dCTP, dGTP, and 400 μ M dUTP, 5 mM MgCl_2 , 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at 50°C and 10 min at 95°C . Thermal cycling consisted of 40 cycles at 95°C for 15 s and 65°C for 1 min. Each assay included: a standard curve (from 10^3 to 10^0 copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

Equipment for real-time detection. The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates C_t and determines the starting copy number in the samples.

Determination of gene amplification. Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

RESULTS

To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erbB2* proto-oncogenes, and the β -amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/ μ l. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as 10^2 copies or as many as 10^5 copies.

Copy-number ratio of the 2 reference genes (*app* and *alb*)

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA

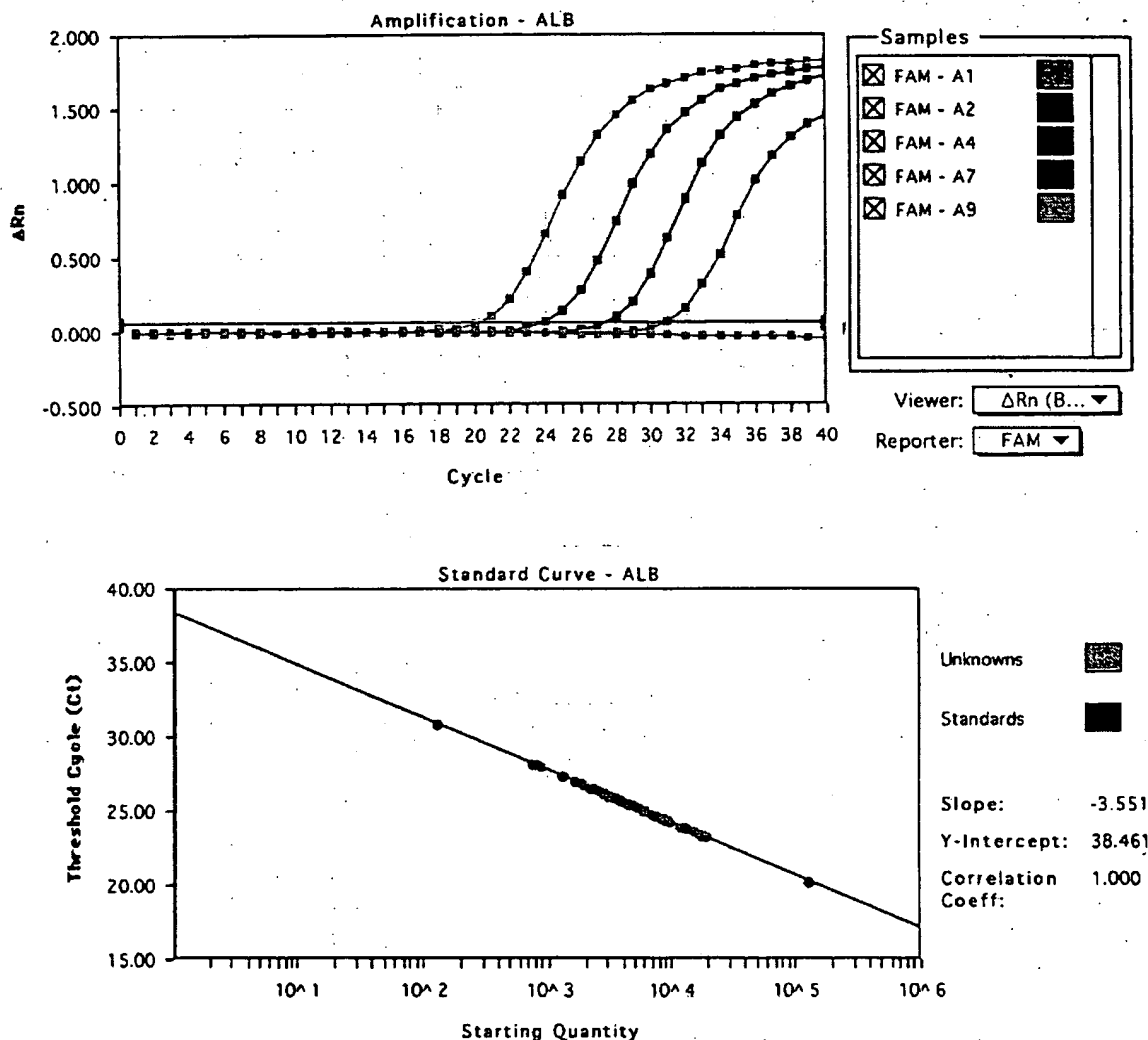


FIGURE 1 – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from 10^5 (A9), 10^4 (A7), 10^3 (A4) to 10^2 (A2) and a no-template control (A1). Cycle number is plotted vs. change in normalized reporter signal (ΔRn). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal (Rn). ΔRn represents the normalized reporter signal (Rn) minus the baseline signal established in the first 15 PCR cycles. ΔRn increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau. C_t (threshold cycle) represents the fractional cycle number at which a significant increase in Rn above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number vs. C_t (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.

samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean 1.02 ± 0.21), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6 , mean 1.06 ± 0.25), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

myc, *ccnd1* and *erbB2* gene dose in normal leukocyte DNA

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean 0.84 ± 0.22) for *myc*; 0.7 to 1.6 (mean 1.06 ± 0.23) for *ccnd1* and 0.6 to 1.3 (mean 0.91 ± 0.19) for *erbB2*. Since N values for *myc*, *ccnd1* and *erbB2* in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

myc, *ccnd1* and *erbB2* gene dose in breast-tumor DNA

myc, *ccnd1* and *erbB2* gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erbB2* (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erbB2*, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (T118). The 3 genes were never found to be co-amplified in the same tumor. *erbB2* and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erbB2* in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erbB2* copy number ($N < 0.5$), suggesting that they bore deletions of the 17q21 region (the site of *erbB2*). No such decrease in copy number was observed with the other 2 proto-oncogenes.

Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erbB2* amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers ($N \geq 5$). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erbB2*) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive

reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

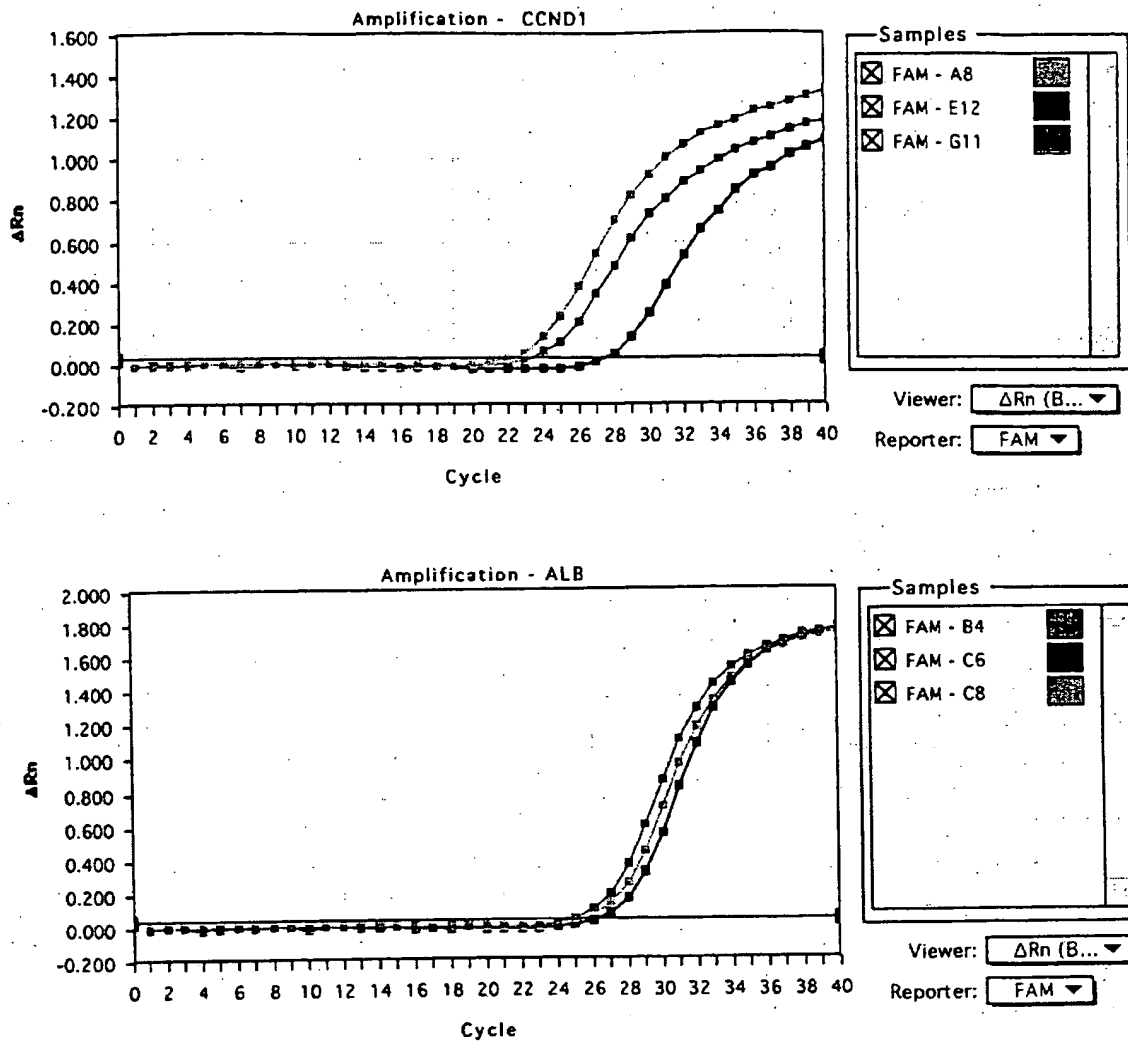
In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on C_t values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI Prism 7700 Sequence Detection System enables C_t to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the C_t value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the C_t ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

The only potential disadvantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that it cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (i) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (ii) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (iii) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erbB2* amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (iv) The maxima of *ccnd1* and *erbB2* over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I - DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erbB2* GENES IN 108 HUMAN BREAST TUMORS

| Gene | Amplification level (N) | | | |
|--------------|-------------------------|------------|------------|----------|
| | <0.5 | 0.5-1.9 | 2-4.9 | ≥ 5 |
| <i>myc</i> | 0 | 97 (89.8%) | 11 (10.2%) | 0 |
| <i>ccnd1</i> | 0 | 83 (76.9%) | 17 (15.7%) | 8 (7.4%) |
| <i>erbB2</i> | 5 (4.6%) | 87 (80.6%) | 8 (7.4%) | 8 (7.4%) |



| Tumor | CCND1 | | ALB | |
|--------|-------|-------------|-------|-------------|
| | C_t | Copy number | C_t | Copy number |
| ■ T118 | 27.3 | 4605 | 26.5 | 4365 |
| ■ T133 | 23.2 | 61659 | 25.2 | 10092 |
| ■ T145 | 22.1 | 125892 | 25.6 | 7762 |

FIGURE 2 - *ccnd1* and *alb* gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the C_t of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (v) The *erbB2* copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron

et al., 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erbB2* amplification in a small series of breast tumors ($n = 25$), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-

TABLE II - EXAMPLES OF *ccnd1* GENE DOSAGE RESULTS FROM 3 BREAST TUMORS¹

| Tumor | <i>ccnd1</i> | | | <i>alb</i> | | | <i>Nccnd1/alb</i> |
|-------|--------------|--------|------|-------------|-------|-----|-------------------|
| | Copy number | Mean | SD | Copy number | Mean | SD | |
| T118 | 4525 | 4603 | 77 | 4223 | 4325 | 89 | 1.06 |
| | 4605 | | | 4365 | | | |
| | 4678 | | | 4387 | | | |
| T133 | 59821 | 61100 | 1111 | 9787 | 10137 | 375 | 6.03 |
| | 61659 | | | 10092 | | | |
| | 61821 | | | 10533 | | | |
| T145 | 128563 | 125392 | 3448 | 7321 | 7672 | 316 | 16.34 |
| | 125892 | | | 7762 | | | |
| | 121722 | | | 7933 | | | |

¹For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (*Nccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers (≥ 5 -fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erbB2*) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra

gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erbB2* (but not of the other 2 proto-oncogenes) in several tumors; *erbB2* is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

REFERENCES

- AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEYEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erbB2* gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, **64**, 291-297 (1995).
- BERNS, E.M.J.J., KLIJN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTINGEN, H. and FOEKENS, J.A., *c-myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res.*, **52**, 1107-1113 (1992).
- BIÈCHE, I. and LIDEREAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, **14**, 227-251 (1995).
- BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., *c-myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, **51**, 687-691 (1992).
- CELLI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDINER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (qd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, **21**, 304-310 (1994).
- COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, **57**, 4360-4367 (1997).
- DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erbB-2* and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, **40**, 271-281 (1996).
- GELMINI, S., ORLANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure *c-erbB-2* oncogene amplification. *Clin. Chem.*, **43**, 752-758 (1997).
- GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, **6**, 995-1001 (1996).
- HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, **6**, 986-994 (1996).
- HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, **88**, 7276-7280 (1991).
- KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, **91**, 2156-2160 (1994).
- LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, **21**, 3761-3766 (1993).
- LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, **93**, 125-128 (1990).
- MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, F., CIRINCIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., *c-erbB-2* expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, **330**, 1260-1266 (1994).
- PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of *HER-2/neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, **13**, 63-72 (1996).
- PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, **14**, 70-80 (1993).
- SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILEE, P., CORNELISSE, C.J., VAN DE VIJVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, **52**, 5229-5234 (1992).
- SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science*, **235**, 177-182 (1987).
- SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDOLV, J., ULLRICH, A. and PRESS, M.F., Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science*, **244**, 707-712 (1989).
- VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying *HER-2/neu* status in breast cancer. *Int. J. Cancer*, **65**, 129-133 (1996).

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft^{‡§}, Thomas Thykjaer[¶], Frederic M. Waldman^{||}, Hans Wolf^{**}, and Julio E. Celis^{‡‡}

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics* 1:37–45, 2002.

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this

From the [‡]Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and ^{**}Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, [¶]AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ^{||}UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ^{‡‡}Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, cyclin d1, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)¹ have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q–, 11p–, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology: examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

¹ The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

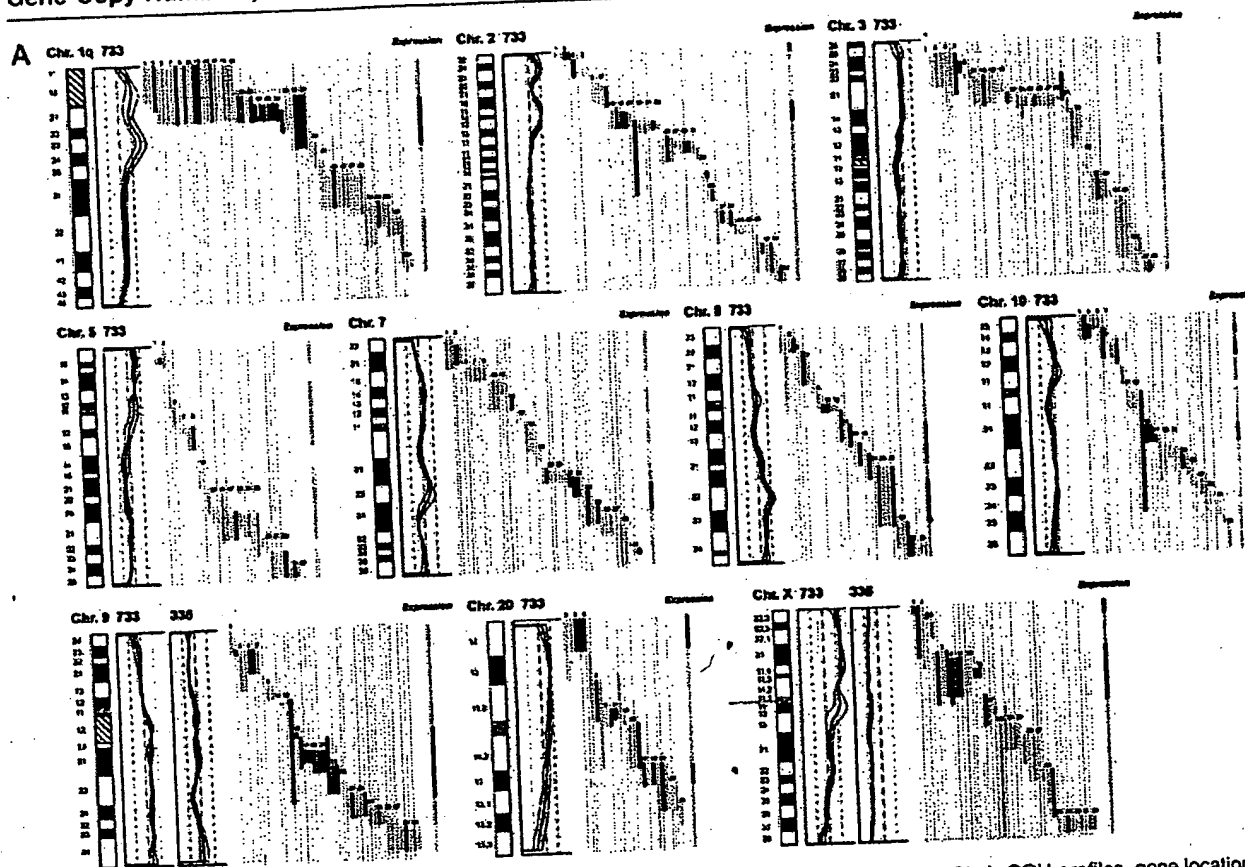


Fig. 1. DNA copy number and mRNA expression level. Shown from left to right are chromosome (Chr.), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. **A**, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. **B**, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (left). The bold curve in the ratio profile represents a mean of four chromosomes and is surrounded by thin curves indicating one standard deviation. The central vertical line (broken) indicates a ratio value of 1 (no change), and the vertical lines next to it (dotted) indicate a ratio of 0.5 (left) and 2.0 (right). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the right of the invasive tumor profile. The colored bars represent one gene each, identified by the running numbers above the bars (the name of the gene can be seen at www.MDL.DK/sdata.html). The bars indicate the purported location of the gene, and the colors indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (black), >2-fold decrease (blue), no significant change (orange). The bar to the far right, entitled *Expression* shows the resulting change in expression along the chromosome; the colors indicate that at least half of the genes were up-regulated (black), at least half of the genes down-regulated (blue), or more than half of the genes are unchanged (orange). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at -80°C . Total RNA was isolated using the RNeasy B RNA isolation method (WAK-Chemie Medical GmbH). poly(A)⁺ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript[®] choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the ME-GAScript[®] *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94°C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6 \times SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95°C for 5 min, subsequently cooled to 40°C , and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40°C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6 \times SSPE-T at 25°C followed by 4 washes in 0.5 \times SSPE-T at 50°C . The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 $\mu\text{g}/\text{ml}$ (Molecular Probes) in 6 \times SSPE-T

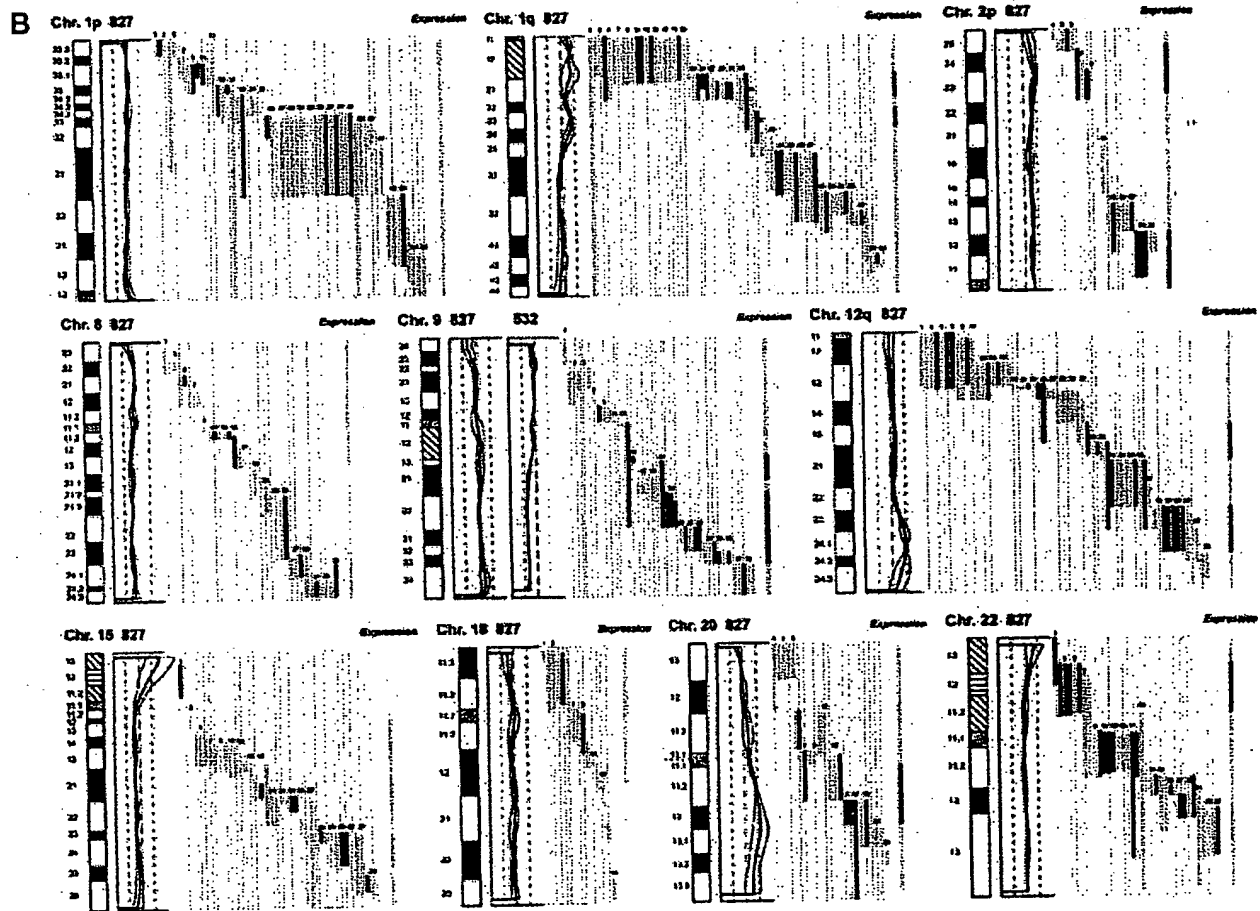


FIG. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μ l for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at -20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/celis.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-

labeled reference DNA (200 ng), and human Cot-1 DNA (20 μ g) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μ g/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

Gene Copy Numbers, Transcripts, and Protein Levels

TABLE I
Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

| independent variable (if expression alteration – what CGH deviation was found). | | | | | | | |
|---|----------------------------|--|-------------|----------------------------|----------------------------|--|-------------|
| CGH alterations | Tumor 733 vs. 335 | | Concordance | CGH alterations | Tumor 827 vs. 532 | | Concordance |
| | Expression change clusters | | | | Expression change clusters | | |
| 13 Gain | 10 Up-regulation | | 77% | 10 Gain | 8 Up-regulation | | 80% |
| | 0 Down-regulation | | | | 0 Down-regulation | | |
| | 3 No change | | | | 2 No change | | |
| 10 Loss | 1 Up-regulation | | 50% | 12 Loss | 3 Up-regulation | | 17% |
| | 5 Down-regulation | | | | 2 Down regulation | | |
| | 4 No change | | | | 7 No change | | |
| Expression change clusters | Tumor 733 vs. 335 | | Concordance | Expression change clusters | Tumor 827 vs. 532 | | Concordance |
| | CGH alterations | | | | CGH alterations | | |
| 16 Up-regulation | 11 Gain | | 69% | 17 Up-regulation | 10 Gain | | 59% |
| | 2 Loss | | | | 5 Loss | | |
| | 3 No change | | | | 2 No change | | |
| 21 Down-regulation | 1 Gain | | 38% | 9 Down-regulation | 0 Gain | | 33% |
| | 8 Loss | | | | 3 Loss | | |
| | 12 No change | | | | 6 No change | | |
| 15 No change | 3 Gain | | 60% | 21 No change | 1 Gain | | 81% |
| | 3 Loss | | | | 3 Loss | | |
| | 9 No change | | | | 17 No change | | |

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22–q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1–qter–, 3q12–q13.3–, 6q12–q22–, 9q34+, 11q12–q13+, 17+, and 20q11.2–q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive versus the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21–q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-

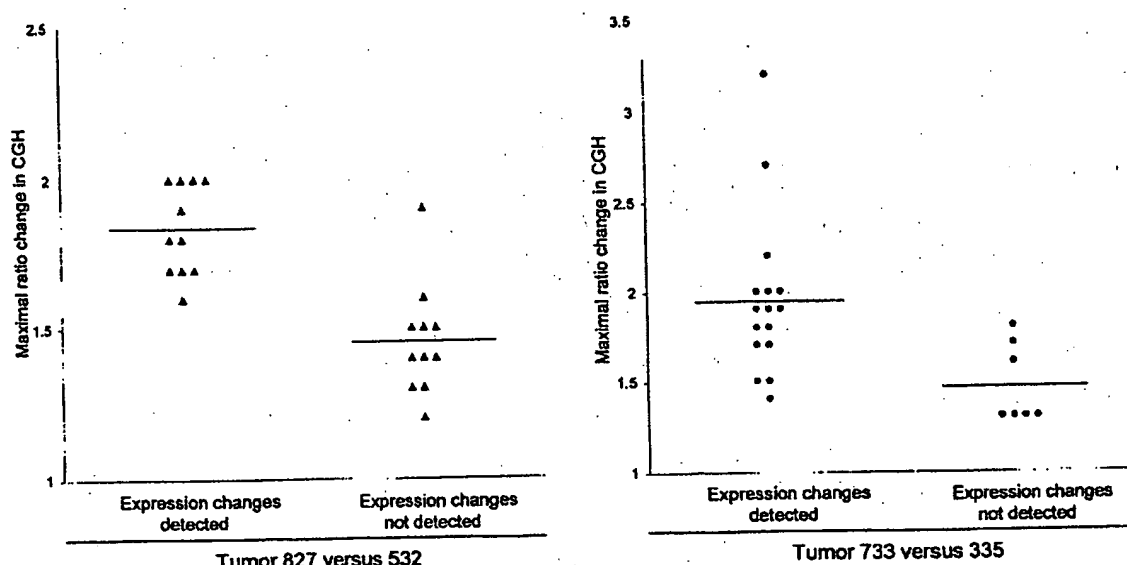


FIG. 2. Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring. The aberration is shown as a numerical fold change in ratio between invasive tumors 827 (Δ) and 733 (\diamond) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the right in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or

because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1A). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

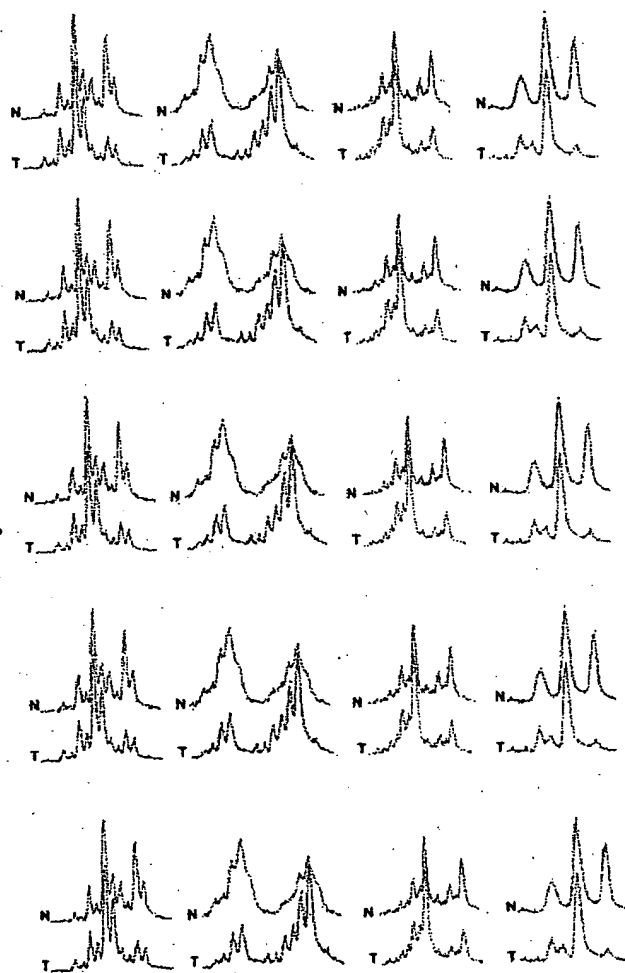


Fig. 3. Microsatellite analysis of loss of heterozygosity. Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (a) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (b) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (c) at chromosome 2p23 by D2S2251 close to general β -spectrin (gene number 11 on Fig. 1) and of (d) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The upper curves show the electropherogram obtained from normal DNA from leukocytes (N), and the lower curves show the electropherogram from tumor DNA (T). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH

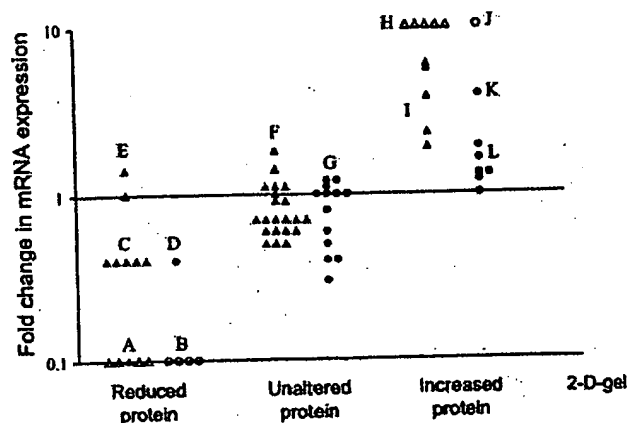


Fig. 4. Correlation between protein levels as judged by 2D-PAGE and transcript ratio. For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (horizontal axis). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (vertical axis). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; △, mRNAs that were scored as absent in the invasive tumors (along horizontal axis) or as absent in non-invasive reference (top of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲▲) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group A (from left), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytochrome P-450, and cytochrome P-450; B (from left), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytochrome P-450, and calyculin; C (from left), α -enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3- ϵ , and pre-mRNA splicing factor; D, mesothelial keratin K7 (type II); E (from top), glutathione S-transferase- π and mesothelial keratin K7 (type II); F (from top and left), adenyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ -actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain- α , hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β -1 subunit; G, (from top and left), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytochrome P-450, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione S-transferase- π , and keratin 8; H (from left), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD $^{+}$ -dependent 15-hydroxyprostaglandin dehydrogenase; I (from top), prolyl 4-hydroxylase β -subunit, cytochrome P-450, keratin 17, prothionin, and fructose 1,6-bisphosphatase; J annexin II; K, annexin IV; L (from top and left), 90-kDa heat shock protein, prolyl 4-hydroxylase β -subunit, α -enolase, GRP 78, cyclophilin, and cofilin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

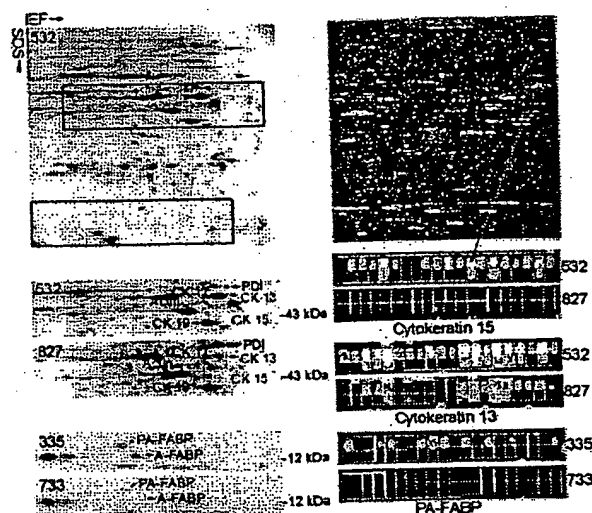


Fig. 5. Comparison of protein and transcript levels in invasive and non-invasive TCCs. The upper part of the figure shows a 2D gel (left) and the oligonucleotide array (right) of TCC 532. The red rectangles on the upper gel highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (red annotation). The tile on the array containing probes for cytokeratin 15 is enlarged below the array (red arrow) from TCC 532 and is compared with TCC 827. The upper row of squares in each tile corresponds to perfect match probes; the lower row corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as black, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the bottom of the figure (left) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the right we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). IEF, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FABP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II
Proteins whose expression level correlates with both mRNA and gene dose changes

| Protein | Chromosomal location | Tumor TCC | CGH alteration | Transcript alteration ^a | Protein alteration |
|-----------------------|----------------------|-----------|----------------|------------------------------------|--------------------|
| Annexin II | 1q21 | 733 | Gain | Abs to Pres ^a | Increase |
| Annexin IV | 2p13 | 733 | Gain | 3.9-Fold up | Increase |
| Cytokeratin 17 | 17q12-q21 | 827 | Gain | 3.8-Fold up | Increase |
| Cytokeratin 20 | 17q21.1 | 827 | Gain | 5.6-Fold up | Increase |
| (PA-)FABP | 8q21.2 | 827 | Loss | 10-Fold down | Decrease |
| FABP1 | 9q22 | 827 | Gain | 2.3-Fold up | Increase |
| Plasma gelsolin | 9q31 | 827 | Gain | Abs to Pres | Increase |
| Heat shock protein 28 | 15q12-q13 | 827 | Loss | 2.5-Fold up | Decrease |
| Prohibitin | 17q21 | 827/733 | Gain | 3.7-/2.5-Fold up ^b | Increase |
| Prolyl-4-hydroxyl | 17q25 | 827/733 | Gain | 5.7-/1.6-Fold up | Increase |
| hnRNPB1 | 7p15 | 827 | Loss | 2.5-Fold down | Decrease |

^a Abs, absent; Pres, present.

^b In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17-19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p-, 9q-, 1q+, Y- (2, 6), and in pT1 tumors, 2q-, 11p-, 11q-, 1q+, 5p+, 8q+, 17q+, and 20q+ (2-4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p- and 9q22-q33- and 9q- and Y-, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22-24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p-, often linked to 20q+ and 11 q13+ (both tumors) (7-9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicates that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Vikolaj Orntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk

REFERENCES

- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* 396, 643–649.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46.
- de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 157–161.
- Brugier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* 12, 1747–1753.
- Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* 50, 1459–1463.
- Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) *c-myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am. J. Pathol.* 146, 1131–1139.
- Richter, J., Jiang, F., Gorog, J. P., Sartorius, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* 57, 2860–2864.
- Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* 153, 1615–1621.
- Bruch, J., Wöhr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Möller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* 23, 167–174.
- Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* 58, 3555–3560.
- Simon, R., Burger, H., Brinkschmidt, C., Bockel, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* 185, 345–351.
- Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* 110, 87–93.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15, 1359–1367.
- Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* 59, 167–177.
- Celis, J. E., Østergaard, M., Basse, B., Cells, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* 56, 4782–4790.
- Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Cells, A. (1994) In *Cell Biology: A Laboratory Handbook* (Cells, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL.
- Ohlsson, R., Tyckö, B., and Sapiezka, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* 14, 435–438.
- Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* 279, 2118–2121.
- Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* 9, 164–170.
- Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmkumpu, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* 59, 3889–3892.
- Cui, H., Hedborg, F., He, L., Nordenskjöld, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 57, 4469–4473.
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* 285, 251–254.
- Tsao, J., Yatabe, Y., Mark, I. D., Haiyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* 29, 26–32.
- Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10632–10636.
- Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537.
- Ideker, T., Thorsson, V., Raniish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934.

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer^{1,2}

Elizabeth Hyman,³ Päivikki Kauraniemi,³ Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkhouloun, Olli-P. Kallioniemi, and Anne Kallioniemi⁴

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E.H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E.H., P.K., S.H., M.W., S.M., E.R., M.R., A.E., O.K., A.K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P.K., A.K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S.H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G.S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O.M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1-6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over

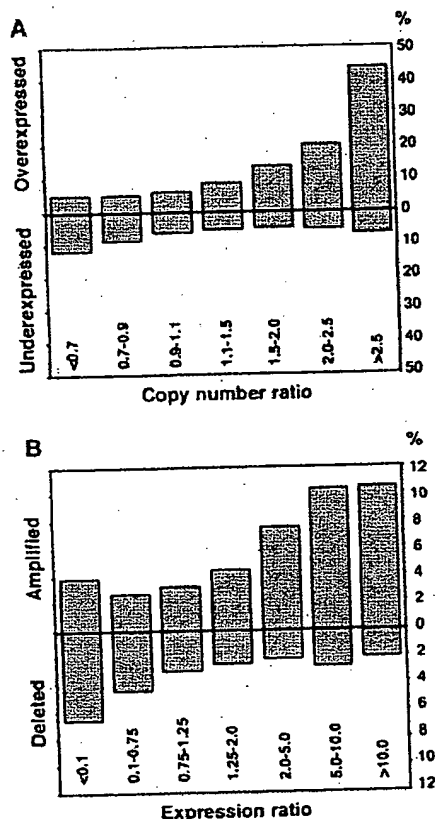


Fig. 1. Impact of gene copy number on global gene expression levels. A, percentage of over- and underexpressed genes (Y axis) according to copy number ratios (X axis). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). B, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7.

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH⁵ (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

² Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

³ Contributed equally to this work.

⁴ To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkeiliijankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

⁵ The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence in situ hybridization; RT-PCR, reverse transcription-PCR.

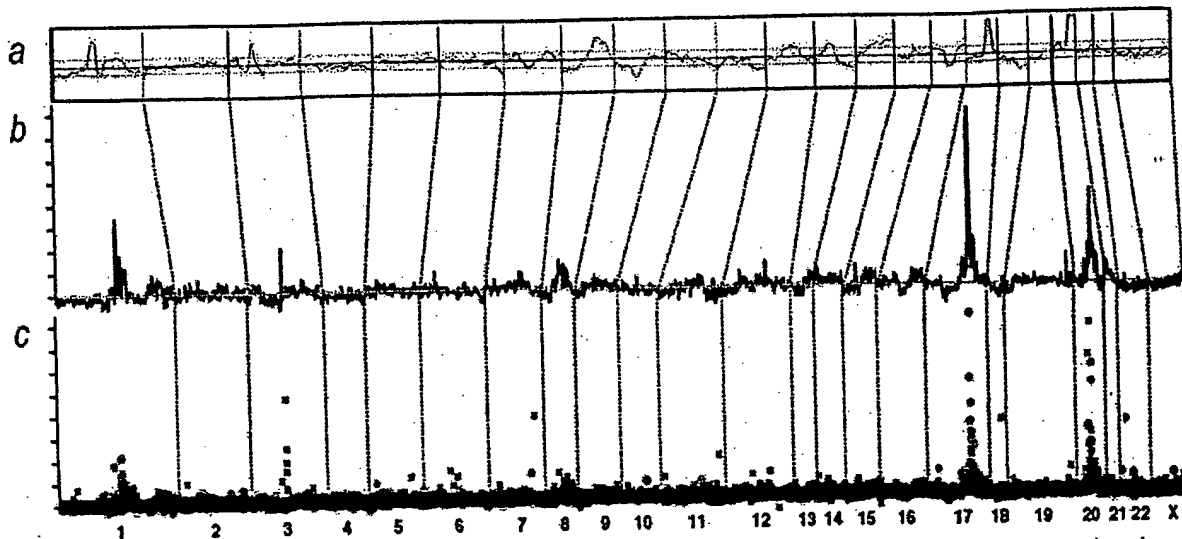


Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (blue line) across the entire genome from 1p telomere to Xq telomere is shown along with ± 1 SD (orange lines). The black horizontal line indicates a ratio of 1.0; red line, a ratio of 0.8; and green line, a ratio of 1.2. *B–C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. Red horizontal line, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The bright red dots indicate the upper 2%, and dark red dots, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); bright green dots indicate the lowest 2%, and dark green dots, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with black crosses. The chromosome numbers are shown at the bottom of the figure, and chromosome boundaries are indicated with a dashed line.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μ g of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *Afl* and *Rsa* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μ g of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μ g of reference RNA were labeled with Cy3-dUTP and 3.5 μ g of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g , for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1} , σ_{g1} , and m_{g0} , σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α . A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.⁶ A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.⁷ The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

⁶ Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.

⁷ Internet address: www.genome.ucsc.edu.

Table 1 Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

| Location | Start (Mb) | End (Mb) | Size (Mb) |
|-----------------|------------|----------|-----------|
| 1p13 | 132.79 | 132.94 | 0.2 |
| 1q21 | 173.92 | 177.25 | 3.3 |
| 1q22 | 179.28 | 179.57 | 0.3 |
| 3p14 | 71.94 | 74.66 | 2.7 |
| 7p12.1-7p11.2 | 55.62 | 60.95 | 5.3 |
| 7q31 | 125.73 | 130.96 | 5.2 |
| 7q32 | 140.01 | 140.68 | 0.7 |
| 8q21.11-8q21.13 | 86.45 | 92.46 | 6.0 |
| 8q21.3 | 98.45 | 103.05 | 4.6 |
| 8q23.3-8q24.14 | 129.88 | 142.15 | 12.3 |
| 8q24.22 | 151.21 | 152.16 | 1.0 |
| 9p13 | 38.65 | 39.25 | 0.6 |
| 13q22-q31 | 77.15 | 81.38 | 4.2 |
| 16q22 | 86.70 | 87.62 | 0.9 |
| 17q11 | 29.30 | 30.85 | 1.6 |
| 17q12-q21.2 | 39.79 | 42.80 | 3.0 |
| 17q21.32-q21.33 | 52.47 | 55.80 | 3.3 |
| 17q22-q23.3 | 63.81 | 69.70 | 5.9 |
| 17q23.3-q24.3 | 69.93 | 74.99 | 5.1 |
| 19q13 | 40.63 | 41.40 | 0.8 |
| 20q11.22 | 34.59 | 35.85 | 1.3 |
| 20q13.12 | 44.00 | 45.62 | 1.6 |
| 20q13.12-q13.13 | 46.45 | 49.43 | 3.0 |
| 20q13.2-q13.32 | 51.32 | 59.12 | 7.8 |

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

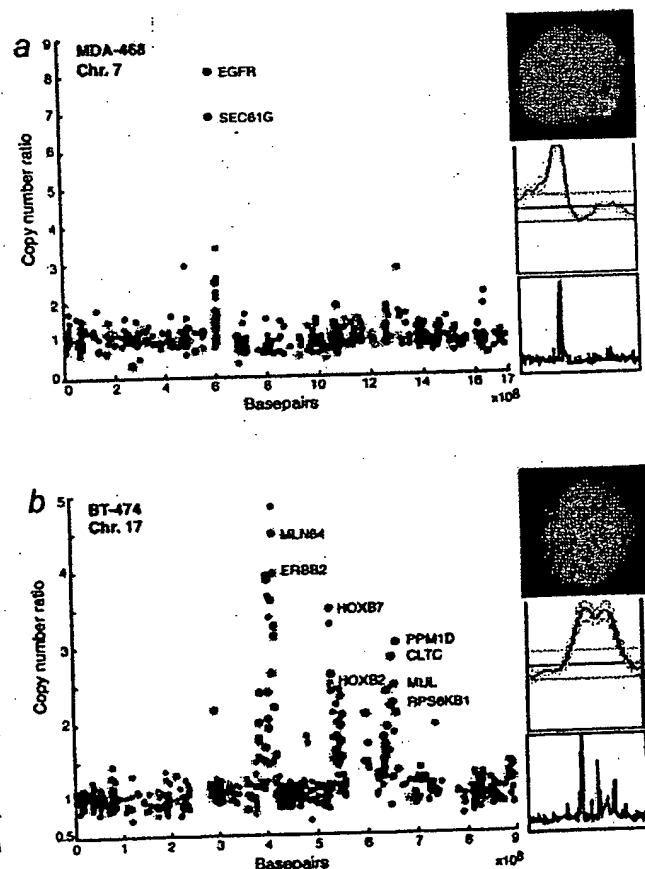


Fig. 3. Annotation of gene expression data on CGH microarray profiles. A, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. B, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).

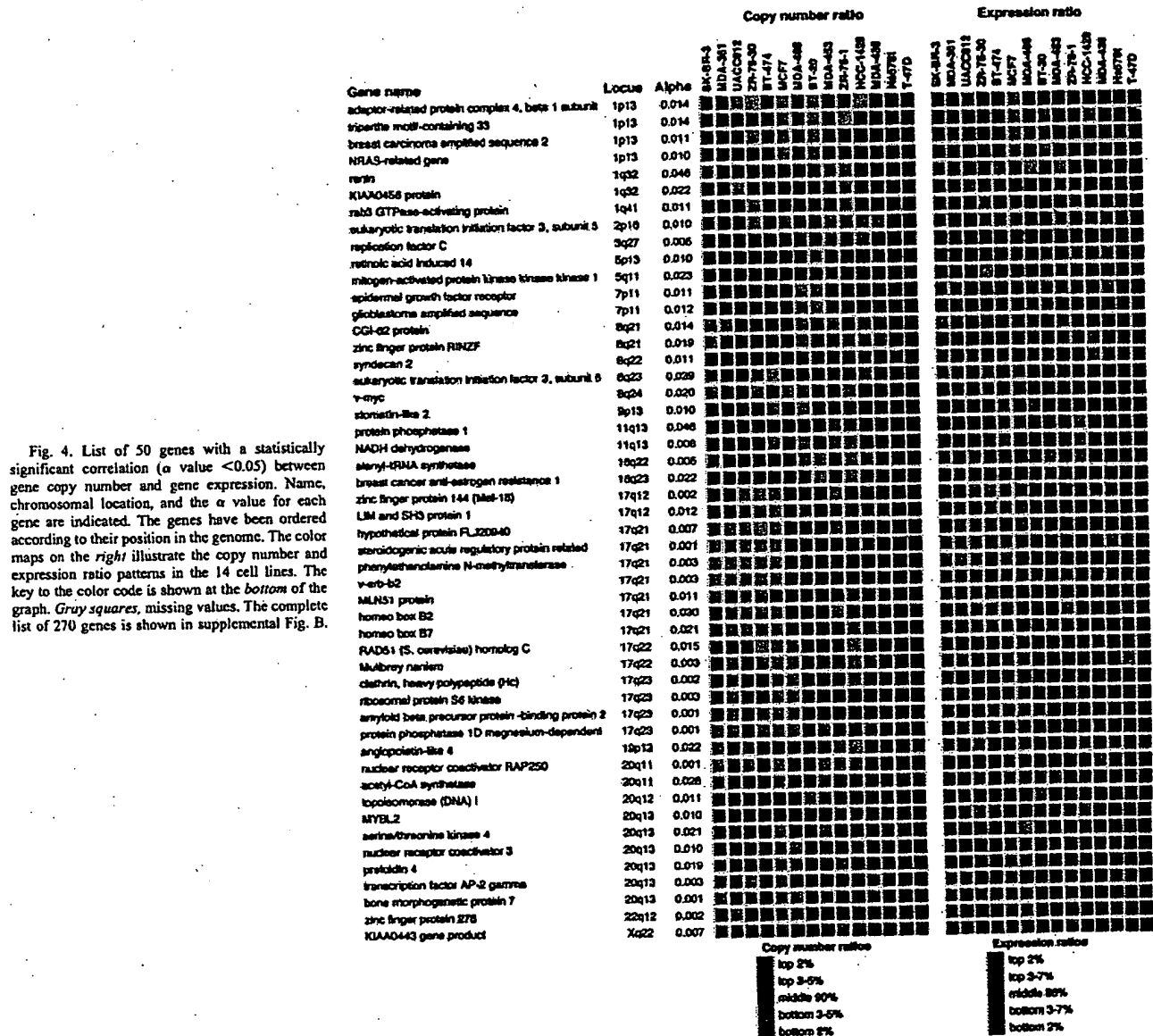


Fig. 4. List of 50 genes with a statistically significant correlation (α value <0.05) between gene copy number and gene expression. Name, chromosomal location, and the α value for each gene are indicated. The genes have been ordered according to their position in the genome. The color maps on the right illustrate the copy number and expression ratio patterns in the 14 cell lines. The key to the color code is shown at the bottom of the graph. Gray squares, missing values. The complete list of 270 genes is shown in supplemental Fig. B.

amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ($P = 0.001$).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,⁸ 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH⁹ (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

⁸ Internet address: <http://www.geneontology.org/>.

⁹ Internet address: <http://www.ncbi.nlm.nih.gov/entrez>.

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22-24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1-2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29-32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein S6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (a) evidence of a prominent global influence of copy number changes on gene expression levels; (b) a high-resolution map of 24 independent amplicons in breast cancer; and (c) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (Wash. DC), 286: 531-537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (Lond.), 403: 503-511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* (Lond.), 406: 536-540, 2000.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. Molecular portraits of human breast tumours. *Nature* (Lond.), 406: 747-752, 2000.
- Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature* (Lond.), 412: 822-826, 2001.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98: 10869-10874, 2001.
- Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, 9: 125-138, 1999.
- Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J. Clin. Oncol.*, 19: 32-40, 2001.
- Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107-1123, 1998.
- Knuutila, S., Autio, K., and Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.*, 157: 689, 2000.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, 14: 457-460, 1996.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, 6: 639-645, 1996.
- Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), *Functional Genomics*, pp. 113-137. Oxford: Oxford University Press, 2000.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, 23: 41-46, 1999.
- Monni, O., Bärklund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711-5716, 2001.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, 2: 364-374, 1997.
- Bärklund, M., Forozan, F., Kononen, J., Eubendorf, L., Chen, Y., Bittner, M. L., Thorhorst, J., Haas, P., Bucher, C., Sauter, G., et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.*, 92: 1252-1259, 2000.
- Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. *Cytometry*, 45: 83-86, 2001.
- Kauraniemi, P., Bärklund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, 61: 8235-8240, 2001.
- Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer*, 34: 104-114, 2002.
- Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res.*, 62: 2625-2629, 2002.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

- Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.*, 25: 333-337, 2000.
23. Viraneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl. Acad. Sci. USA*, 98: 1124-1129, 2001.
 24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linchan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.*, 61: 8143-8149, 2001.
 25. Bärthel, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer*, 20: 372-376, 1997.
 26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsynthetic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.*, 56: 3441-3445, 1996.
 27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. *Exp. Cell Res.*, 248: 1-9, 1999.
 28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.*, 188: 161-169, 2001.
 29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell. Biol.*, 16: 4842-4851, 1996.
 30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene*, 16: 3285-3289, 1998.
 31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res.*, 61: 6532-6539, 2001.
 32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. *Proc. Natl. Acad. Sci. USA*, 98: 4060-4065, 2001.

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack^{*,†,‡}, Therese Sørlie[§], Charles M. Perou[¶], Christian A. Rees^{||}, Stefanie S. Jeffrey^{††}, Per E. Lonning^{††}, Robert Tibshirani^{§§}, David Botstein^{||}, Anne-Lise Børresen-Dale[§], and Patrick O. Brown^{†††}

Departments of ^{*}Pathology, [†]Genetics, ^{††}Surgery, ^{§§}Health Research and Policy, and ^{¶¶}Biochemistry, and ^{†††}Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; [§]Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ^{||}Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ^{††}Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From

this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. "Test" DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The "reference" (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at <http://rana.lbl.gov>). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

^{*}To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

^{††}Present address: Zymoix Inc., Hayward, CA 94545.

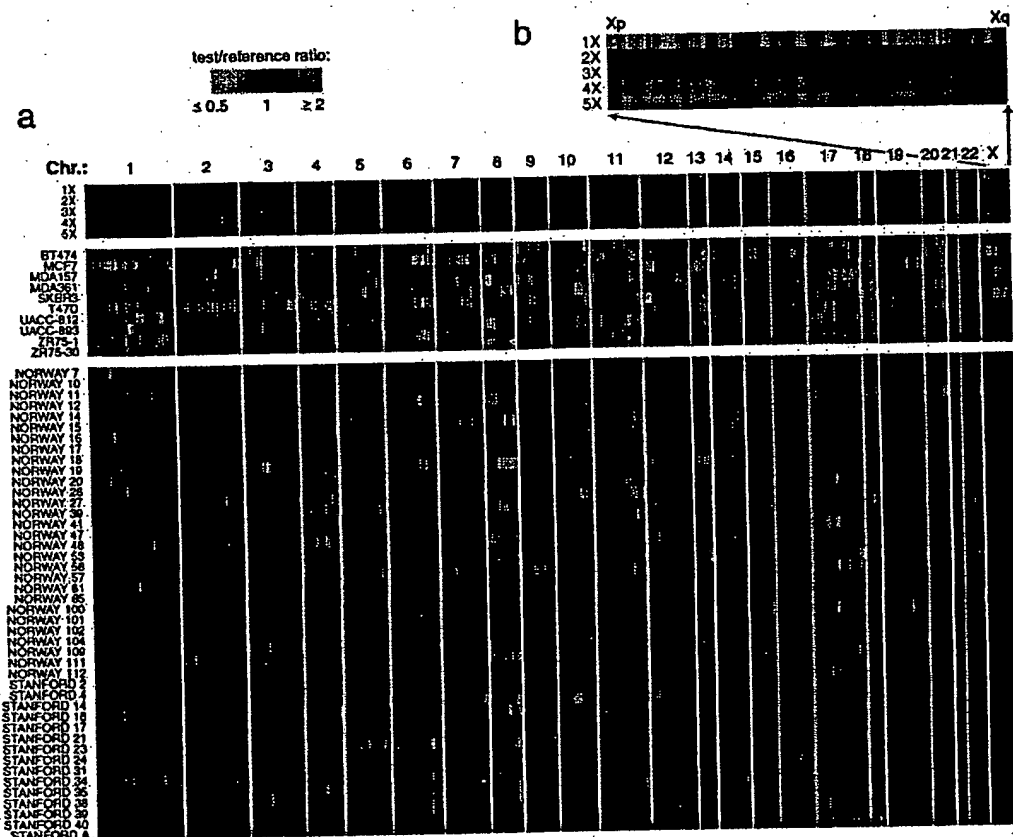


Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a \log_2 -based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (b) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (<http://genome.ucsc.edu/>; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1a; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (<http://genome.ucsc.edu/>) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and

deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1b), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1a), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2-4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total

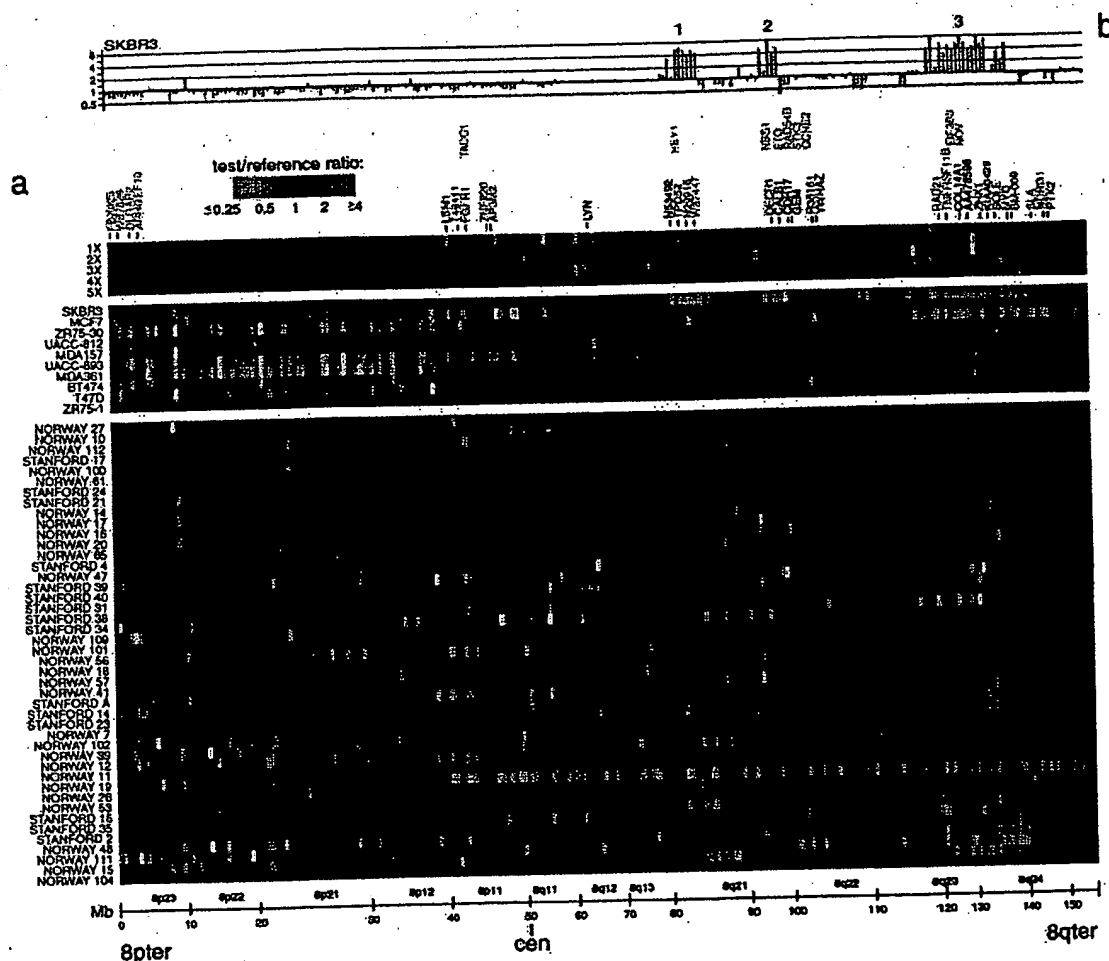


Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a log₂ pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (b) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a log₂ scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2a). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2b). For each of these regions we can define the

boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

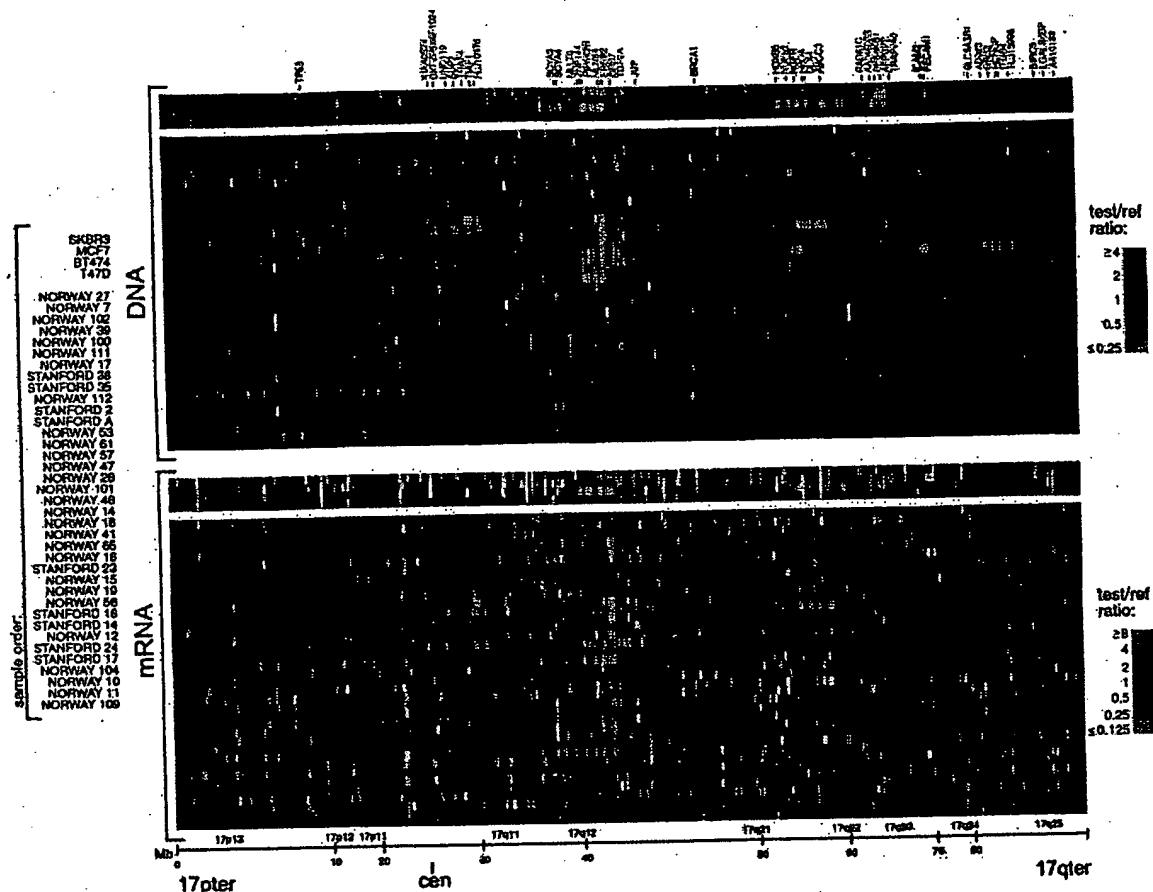


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (Upper) and mRNA levels (Lower) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (Upper), and the identical sample order is maintained (Lower). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate \log_2 pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4 , and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4a). For both the

breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (P values for pair-wise Student's t tests comparing adjacent classes: cell lines, 4×10^{-49} , 1×10^{-49} , 5×10^{-5} , 1×10^{-2} ; tumors, 1×10^{-43} , 1×10^{-214} , 5×10^{-41} , 1×10^{-4}). A linear regression of the average $\log(\text{DNA copy number})$, for each class, against average $\log(\text{mRNA level})$ demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4a, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4b). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4c), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4b) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37

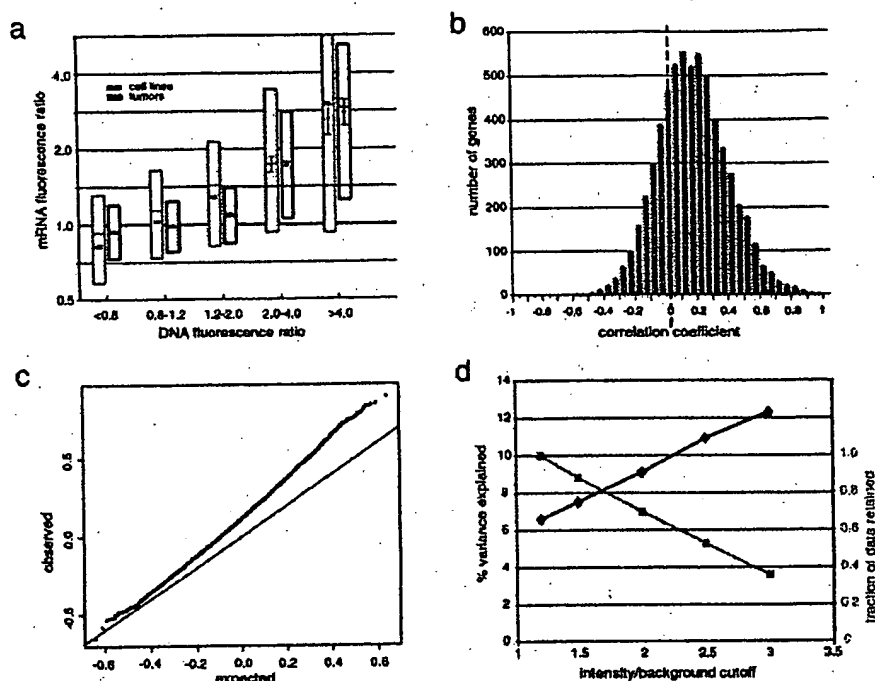


Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (a) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log₂ scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49} , 1×10^{-49} , 5×10^{-5} , 1×10^{-2} (cell lines), and 1×10^{-43} , 1×10^{-214} , 5×10^{-41} , 1×10^{-4} (tumors). (b) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (c) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (d) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4d). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background > 3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4d). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer

cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousands residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing

the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stoichiometric relationships in cell metabolism and physiology (e.g., proteasome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818–821.
2. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156–2160.
3. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177–184.
4. Forozan, F., Mahlamaki, E. H., Mooni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519–4525.
5. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 20, 399–407.
6. Pinkel, D., Segreaves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207–211.
7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41–46.
8. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747–752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869–10874.
10. Schuler, G. D. (1997) *J. Mol. Med.* 75, 694–698.
11. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860–921.
12. Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105–113.
13. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333–337.
14. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143–8149.
15. Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Wilton, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134–1138.
16. Albertson, D. G., Ylstra, B., Segreaves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144–146.
17. Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506–14511.
18. Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 340, 621–630.



TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

HER-2/neu Breast Cancer Predictive Testing

Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.¹ Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.²

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.³

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in

recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.⁴ HER-2/neu status may be particularly important to establish in women with small (≤ 1 cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.⁵ Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin® (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin® in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin® in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest®) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

CPT code information

HER-2/neu via IHC

88342 (including interpretive report)

HER-2/neu via FISH

- 88271x2 Molecular cytogenetics, DNA probe, each
88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells
88291 Cytogenetics and molecular cytogenetics, interpretation and report

Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest[®]. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion[™] HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The PathVysion[™] kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2. "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-2/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F. Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

Provided for the clients of

**PATHOLOGY ASSOCIATES MEDICAL LABORATORIES
PACLAB NETWORK LABORATORIES
TRI-CITIES LABORATORY
TREASURE VALLEY LABORATORY**

*For more information, please contact
your local representative.*

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE||, ROBERT L. COHEN*, MONA F. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,‡‡, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, ‡Molecular Biology, §Scientific Computing, and ¶Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 β (GSK-3 β) resulting in an increase in β -catenin levels. Stabilized β -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β -catenin levels (9). APC is phosphorylated by GSK-3 β , binds to β -catenin, and facilitates its degradation. Mutations in either APC or β -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- β superfamily, and the homeobox genes, *engrailed*, *gooseoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0
PNAS is available online at www.pnas.org.

cDNA was synthesized from 2 μ g of poly(A)⁺ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μ g of poly(A)⁺ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

In Situ Hybridization. ³³P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNazol.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{(\Delta Ct)}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The Δ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1 *A* and *B*). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of \sim 40,000 (M_r , 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2*A*).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of \sim 27,000 (M_r , 27 K) (Fig. 2*B*). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

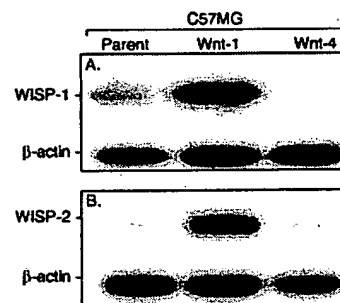


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (*A*) and *WISP-2* (*B*) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)⁺ RNA (2 μ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human β -actin probe.

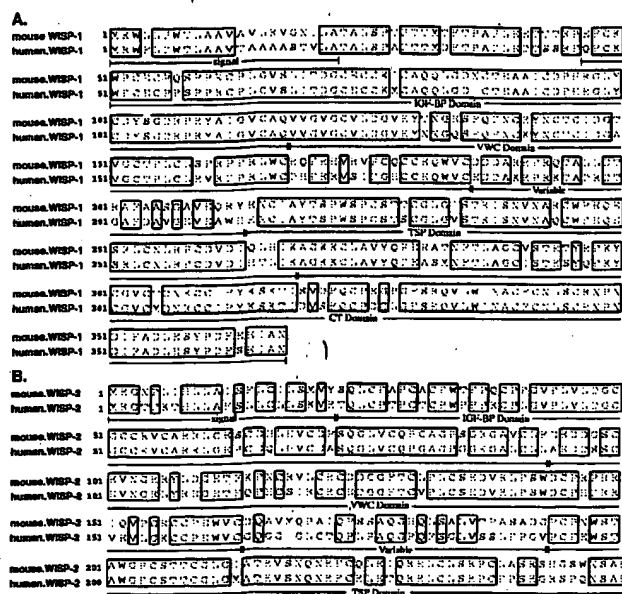


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

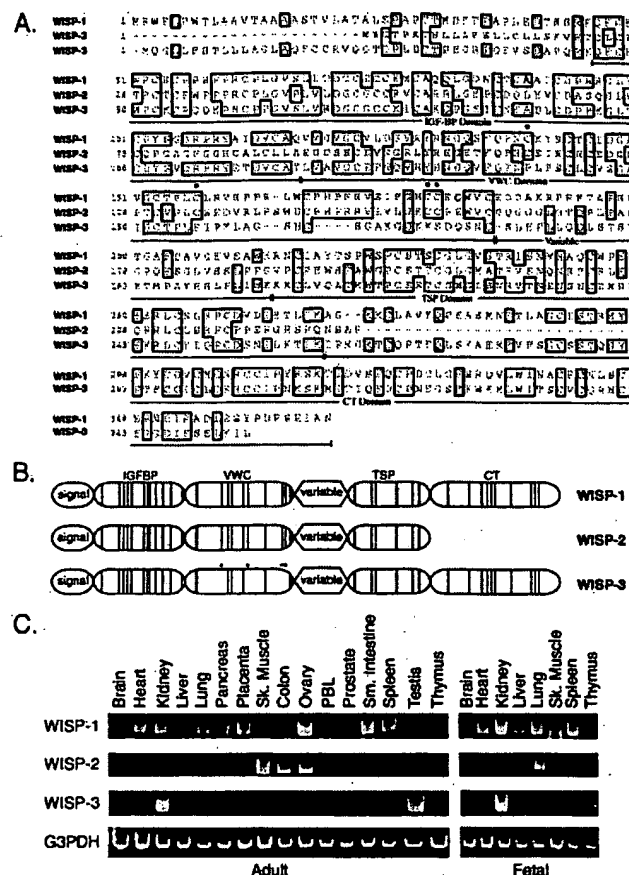


FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSXCSXCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

In Situ Localization of *WISP-1* and *WISP-2*. Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

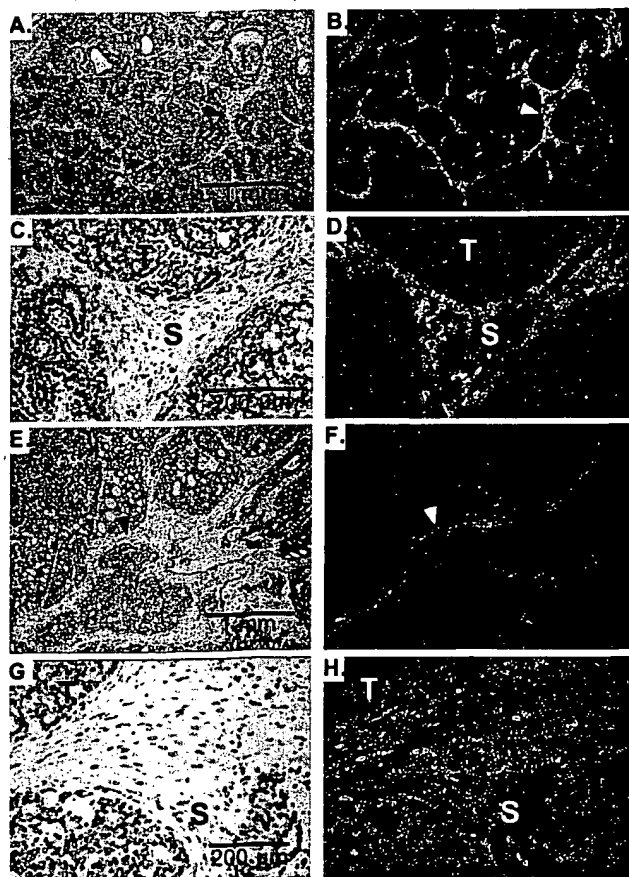


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

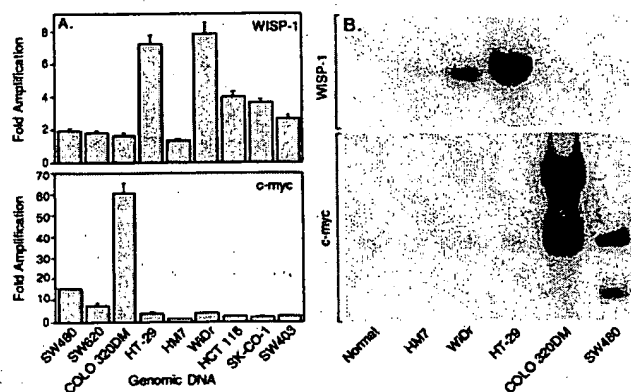


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10 μ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

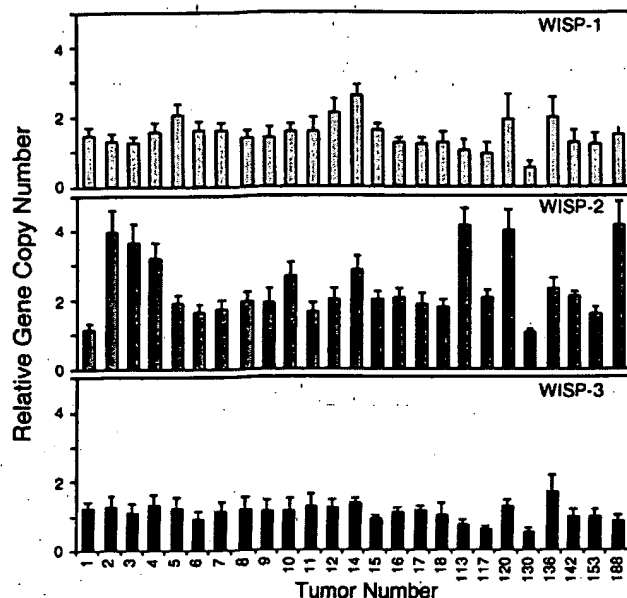


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

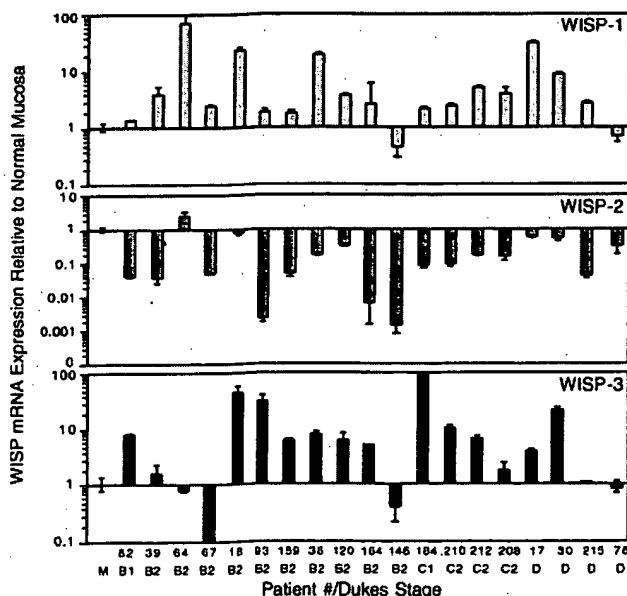


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa by the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- β , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- β 1, which is the stimulus for stromal proliferation (34). TGF- β 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12-20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36-38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISPs*. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISPs* as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISPs* in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286-3305.
- Dale, T. C. (1998) *Biochem. J.* 329, 209-223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99-109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233-240.
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619-625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182-187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391-399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248-1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046-3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509-1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025-6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001-1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278-6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349-1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289-296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131-6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171-179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326-1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355-6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729-2732.
- Bork, P. (1993) *FEBS Lett.* 327, 125-130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981-12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiat, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10-21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514-19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852-2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195-205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529-2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell Genet.* 57, 109-111.
- Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88-94.
- Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435-449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268-1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269-291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090-3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153-161.
- Wernert, N. (1997) *Virchows Arch.* 430, 433-443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., et al. (1994) *Cancer Res.* 54, 4257-4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187-194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., et al. (1998) *EMBO J.* 17, 3052-3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787-1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169-176.

Aneuploidy and cancer

Subrata Sen, PhD

Numeric aberrations in chromosomes, referred to as aneuploidy, is commonly observed in human cancer. Whether aneuploidy is a cause or consequence of cancer has long been debated. Three lines of evidence now make a compelling case for aneuploidy being a discrete chromosome mutation event that contributes to malignant transformation and progression process. First, precise assay of chromosome aneuploidy in several primary tumors with *in situ* hybridization and comparative genomic hybridization techniques have revealed that specific chromosome aneusomies correlate with distinct tumor phenotypes. Second, aneuploid tumor cell lines and *in vitro* transformed rodent cells have been reported to display an elevated rate of chromosome instability, thereby indicating that aneuploidy is a dynamic chromosome mutation event associated with transformation of cells. Third, and most important, a number of mitotic genes regulating chromosome segregation have been found mutated in human cancer cells, implicating such mutations in induction of aneuploidy in tumors. Some of these gene mutations, possibly allowing unequal segregations of chromosomes, also cause tumorigenic transformation of cells *in vitro*. In this review, the recent publications investigating aneuploidy in human cancers, rate of chromosome instability in aneuploidy tumor cells, and genes implicated in regulating chromosome segregation found mutated in cancer cells are discussed. Curr Opin Oncol 2000, 12:82-88 © 2000 Lippincott Williams & Wilkins, Inc.

The University of Texas, M.D. Anderson Cancer Center, Department of Laboratory Medicine, Houston, Texas, USA

Correspondence to Subrata Sen, PhD, The University of Texas, MD Anderson Cancer Center, Department of Laboratory Medicine, Box 054, 1515 Holcombe Blvd., Houston, TX 77030, USA; tel: 713-792-2560; fax: 713-792-4094; e-mail: ssen@mdanderson.org

Current Opinion in Oncology 2000 12:82-88

Abbreviations

CGH comparative genomic hybridization
CHE Chinese hamster embryo cells
FISH fluorescence *in situ* hybridization
HPRC hereditary papillary renal carcinoma
ISH *in situ* hybridization

ISSN 1040-8746 © 2000 Lippincott Williams & Wilkins, Inc.

Cancer research over the past decade has firmly established that malignant cells accumulate a large number of genetic mutations that affect differentiation, proliferation, and cell death processes. In addition, it is also recognized that most cancers are clonal, although they display extensive heterogeneity with respect to karyotypes and phenotypes of individual clonal populations. It is estimated that numeric chromosomal imbalance, referred to as *aneuploidy*, is the most prevalent genetic change recorded among over 20,000 solid tumors analyzed thus far [1]. Phenotypic diversity of the clonal populations in individual tumors involve differences in morphology, proliferative properties, antigen expression, drug sensitivity, and metastatic potentials. It has been proposed that an underlying acquired genetic instability is responsible for the multiple mutations detected in cancer cells that lead to tumor heterogeneity and progression [2]. In a somewhat contradictory argument, it has also been suggested that clonal expansion due to selection of cells undergoing normal rates of mutation can explain malignant transformation and progression process in humans [3]. Acquired genetic instability, nonetheless, is considered important for more rapid progression of the disease [4••]. Although the original hypothesis on genetic instability in cancer primarily focused on chromosome imbalances in the form of aneuploidy in tumor cells, the actual relevance of such mutations in cancer remains a controversial issue.

Whether or not aneuploidy contributes to the malignant transformation and progression process has long been debated. A prevalent idea on genetics of cancer referred to as "somatic gene mutation hypothesis" contends that gene mutations at the nucleotide level alone can cause cancer by either activating cellular proto-oncogenes to dominant cancer causing oncogenes and/or by inactivating growth inhibitory tumor suppressor genes. In this scheme of things chromosomal instability in the form of aneuploidy is a mere consequence rather than a cause of malignant transformation and progression process.

In this review, some of the recent observations on the subject are discussed and compelling evidence is provided to suggest that aneuploidy is a distinct form of genetic instability in cancer that frequently correlates with specific phenotypes and stages of the disease. Furthermore, discrete genetic targets affecting chromosomal stability in cancer cells, recently identified, are also discussed. These data provide a new direction toward elucidating the molecular mechanisms responsi-

ble for induction of aneuploidy in cancer and may eventually be exploited as novel therapeutic targets in the future.

Genetic alterations in cancer

Alterations in many genetic loci regulating growth, senescence, and apoptosis, identified in tumor cells, have led to the current understanding of cancer as a genetic disease. The genetic changes identified in tumors include: subtle mutations in genes at the nucleotide level; chromosomal translocations leading to structural rearrangements in genes; and numeric changes in either partial segments of chromosomes or whole chromosomes (aneuploidy) causing imbalance in gene dosage.

For the purpose of this review, both segmental and whole chromosome imbalances leading to altered DNA dosage in cancer cells are included as examples of aneuploidy.

Incidence of aneuploidy in cancer

Evidence of aneuploidy involving one or more chromosomes have been commonly reported in human tumors. Although these observations were initially made using classic cytogenetic techniques late in a tumor's evolution and were difficult to correlate with cancer progression, more recent studies have reported association of specific nonrandom chromosome aneuploidy with different biologic properties such as loss of hormone dependence and metastatic potential [5].

Classic cytogenetic studies performed on tumor cells had serious limitations in scope because they were applicable only to those cases in which mitotic chromosomes could be obtained. Because of low spontaneous rates of cell division in primary tumors, analyses depended on cells either derived selectively from advanced metastases or those grown *in vitro* for variable periods of time. In both instances, metaphases analyzed represented only a subset of primary tumor cell population. Two major advances in cytogenetic analytic techniques, *in situ* hybridization (ISH) and comparative genomic hybridization (CGH), have allowed better resolution of chromosomal aberrations in freshly isolated tumor cells [6]. ISH analyses with chromosome-specific DNA probes, a powerful adjunct to metaphasic analysis, allows assessment of chromosomal anomalies within tumor cell populations in the contexts of whole nuclear architecture and tissue organization. CGH allows genome wide screening of chromosomal anomalies without the use of specific probes even in the absence of prior knowledge of chromosomes involved. Although both techniques have certain limitations in terms of their resolution power, they nonetheless provide a better approximation of chromosomal changes occurring among tumors of various histology, grade, and stage

compared with what was possible with the classic cytogenetic techniques. Genomic ploidy measurements have also been performed at the DNA level with flow cytometry and cytofluorometric methods. Although these assays underestimate chromosome ploidy due to a chromosomal gain occasionally masking a chromosomal loss in the same cell, several studies using these methods have supported the conclusion that DNA aneuploidy closely associates with poor prognosis in various cancers [7,8]. This discussion of some recent examples published on aneuploidy in cancer includes discussion of studies dealing with DNA ploidy measurements as well. Most of these observations are correlative without direct proof of specific involvement of genes on the respective chromosomes. Identification of putative oncogenes and tumor suppressor genes on gained and lost chromosomes in aneuploid tumors, however, are providing strong evidence that chromosomes involved in aneuploidy play a critical role in the tumorigenic process.

In renal tumors, either segmental or whole chromosome aneuploidy appears to be uniquely associated with specific histologic subtypes [9]. Tumors from patients with hereditary papillary renal carcinomas (HPRC) commonly show trisomy of chromosome 7, when analyzed by CGH. Germline mutations of a putative oncogene *MET* have been detected in patients with HPRC. A recent study [10] has demonstrated that an extra copy of chromosome 7 results in nonrandom duplication of the mutant *MET* allele in HPRC, thereby implicating this trisomy in tumorigenesis. The study suggested that mutation of *MET* may render the cells more susceptible to errors in chromosome replication, and that clonal expansion of cells harboring duplicated chromosome 7 reflects their proliferative advantage. In addition to chromosome 7, trisomy of chromosome 17 in papillary tumors and also of chromosome 8 in mesoblastic nephroma are commonly seen. Association of specific chromosome imbalances with benign and malignant forms of papillary renal tumors, therefore, not only contribute to an understanding of tumor origins and evolution, but also implicate aneuploidy of the respective chromosomes in the tumorigenic transformation process.

In colorectal tumors, chromosome aneuploidy is a common occurrence. In fact, molecular allelotyping studies have suggested that limited karyotyping data available from these tumors actually underestimate the true extent of these changes. Losses of heterozygosity reflecting loss of the maternal or paternal allele in tumors are widespread and often accompanied by a gain of the opposite allele. Therefore, for example, a tumor could lose a maternal chromosome while duplicating the same paternal chromosome, leaving the tumor cell

with a normal karyotype and ploidy but an aberrant allelotype. It has been estimated that cancer of the colon, breast, pancreas, or prostate may lose an average of 25% of its alleles. It is not unusual to discover that a tumor has lost over half of its alleles [4]. In clinical settings, DNA ploidy measurements have revealed that DNA aneuploidy indicates high risk of developing severe premalignant changes in patients with ulcerative colitis, who are known to have an increased risk of developing colorectal cancer [11]. DNA aneuploidy has been found to be one of the useful indicators of lymph node metastasis in patients with gastric carcinoma and associated with poor outcome compared with diploid cases [12,13]. CGH analyses of chromosome aneuploidy, on the other hand, was reported to correlate gain of chromosome 20q with high tumor S phase fractions and loss of 4q with low tumor apoptotic indices [14]. Aneuploidy of chromosome 4 in metastatic colorectal cancer has recently been confirmed in studies that used unbiased DNA fingerprinting with arbitrarily primed polymerase chain reactions to detect moderate gains and losses of specific chromosomal DNA sequences [15]. The molecular karyotype (amplotype) generated from colorectal cancer revealed that moderate gains of sequences from chromosomes 8 and 13 occurred in most tumors, suggesting that overrepresentation of these chromosomal regions is a critical step for metastatic colorectal cancer.

In addition to being implicated in tumorigenesis and correlated with distinct tumor phenotypes, chromosome aneuploidy has been used as a marker of risk assessment and prognosis in several other cancers. The potential value of aneuploidy as a noninvasive tool to identify individuals at high risk of developing head and neck cancer appears especially promising. Interphase fluorescence *in situ* hybridization (FISH) revealed extensive aneuploidy in tumors from patients with head and neck squamous cell carcinomas (HNSCC) and also in clinically normal distant oral regions from the same individuals [16,17]. It has been proposed that a panel of chromosome probes in FISH analyses may serve as an important tool to detect subclinical tumorigenesis and for diagnosis of residual disease. The presence of aneuploid or tetraploid populations is seen in 90% to 95% of esophageal adenocarcinomas, and when seen in conjunction with Barrett's esophagus, a premalignant condition, predicts progression of disease [18,19]. Chromosome ploidy analyses in conjunction with loss of heterozygosity and gene mutation studies in Barrett's esophagus reflect evolution of neoplastic cell lineages *in vivo* [20]. Evolution of neoplastic progeny from Barrett's esophagus following somatic genetic mutations frequently involves bifurcations and loss of heterozygosity at several chromosomal loci leading to aneuploidy and cancer. Accordingly, it is hypothesized that during

tumor cell evolution diploid cell progenitors with somatic genetic abnormalities undergo expansion with acquired genetic instability. Such instability, often manifested in the form of increased incidence of aneuploidy, enters a phase of clonal evolution beginning in premalignant cells that proceeds over a period of time and occasionally leads to malignant transformation. The clonal evolution continues even after the emergence of cancer.

The significance of DNA and chromosome aneuploidy in other human cancers continue to be evaluated. Among papillary thyroid carcinomas, aneuploid DNA content in tumor cells was reported to correlate with distant metastases, reflecting worsened prognosis [21]. Genome wide screening of follicular thyroid tumors by CGH, on the other hand, revealed frequent loss of chromosome 22 in widely invasive follicular carcinomas [22]. Chromosome copy number gains in invasive neoplasm compared with foci of ductal carcinoma *in situ* (DCIS) with similar histology have been proposed to indicate involvement of aneuploidy in progression of human breast cancer [23]. ISH analyses of cervical intraepithelial neoplasia has provided suggestive evidence that chromosomes 1, 7 and X aneusomy is associated with progression toward cervical carcinoma [24].

Although the prognostic value of numeric aberrations remains a matter of debate in human hematopoietic neoplasia, there have been recent studies to suggest that the presence of monosomy 7 defines a distinct subgroup of acute myeloid leukemia patients [25]. It is interesting in this context that therapy-related myelodysplastic syndromes have been reported to display monosomy 5 and 7 karyotypes, reflecting poor prognosis [26].

The clinical observations, mentioned previously, are supported by *in vitro* studies in human and rodent cells in which aneuploidy is induced at early stages of transformation [27,28]. It is even suggested that aneuploidy may cause cell immortalization, in some instances, that is a critical step preceding transformation.

Finally, in an interesting study to develop transgenic mouse models of human chromosomal diseases, chromosome segment specific duplication and deletions of the genome were reported to be constructed in mouse embryonic stem cells [29]. Three duplications for a portion of mouse chromosome 11 syntenic with human chromosome 17 were established in the mouse germline. Mice with 1Mb duplication developed corneal hyperplasia and thymic tumors. The findings represent the first transgenic mouse model of aneuploidy of a defined chromosome segment that documents the direct role of chromosome aneusomy in tumorigenesis.

Aneuploidy as "dynamic cancer-causing mutation" instead of a "consequential state" in cancer

According to the hypothesis previously discussed, aneuploidy represents either a "gain of function" or "loss of function" mutation at the chromosome level with a causative influence on the tumorigenesis process. The hypothesis, however, is based only on circumstantial evidence even though existence of aneuploidy is correlated with different tumor phenotypes. The existence of numeric chromosomal alterations in a tumor does not mean that the change arose as a dynamic mutation due to genomic instability, because several factors could lead to consequential aneuploidy in tumors, also. Although aneuploidy as a dynamic mutation due to genomic instability in tumor cells would occur at a certain measurable rate per cell generation, a consequential state of aneuploidy in tumors may not occur at a predictable rate under similar conditions or in tumors with similar phenotypes. In addition to genomic instability, differences in environmental factors with selective pressure, could explain high incidence of aneuploidy and other somatic mutations in tumors compared with normal cells [4]. These include humoral, cell substratum, and cell-cell interaction differences between tumor and normal cell environments. It could be argued that despite similar rates of spontaneous aneuploidy induction in normal and tumor cells, the latter are selected to proliferate due to altered selective pressure in the tumor cell environment, whereas the normal cells are eliminated through activation of apoptosis. Alternatively, of course, one could postulate that selective expression or overexpression of anti-apoptotic proteins or inactivation of proapoptotic proteins in tumor cells may counteract default induction of apoptosis in G2/M phase cells undergoing missegregation of chromosomes. Recent demonstration of overexpression of a G2/M phase anti-apoptotic protein survivin in cancer cells [30] suggests that this protein may favor aberrant progression of aneuploid transformed cells through mitosis. This would then lead to proliferation of aneuploid cell lineages, which may undergo clonal evolution.

To ascertain that aneuploidy is a dynamic mutational event, various human tumor cell lines and transformed rodent cell lines have been analyzed for the rate of aneuploidy induction. When grown under controlled *in vitro* conditions, such conditions ensure that environmental factors do not influence selective proliferation of cells with chromosome instability. In one study, Lengauer *et al.* [31•] provided unequivocal evidence by FISH analyses that losses or gains of multiple chromosomes occurred in excess of 10^{-2} per chromosome per generation in aneuploid colorectal cancer cell lines. The study further concluded that such chromosomal instability appeared to be a dominant trait. Using another *in*

vitro model system of Chinese hamster embryo (CHE) cells, Duesberg *et al.* [32•] have also obtained similar results. With clonal cultures of CHE cells, transformed with nongenotoxic chemicals and a mitotic inhibitor, these authors demonstrated that the overwhelming majority of the transformed colonies contained more than 50% aneuploid cells, indicating that aneuploidy would have originated from the same cells that underwent transformation. All the transformed colonies tested were tumorigenic. It was further documented that the ploidy factor representing the quotient of the modal chromosome number divided by the normal diploid number, in each clone, correlated directly with the degree of chromosomal instability. Therefore, chromosomal instability was found proportional to the degree of aneuploidy in the transformed cells and the authors hypothesized that aneuploidy is a unique mechanism of simultaneously altering and destabilizing, in a massive manner, the normal cellular phenotypes. In the absence of any evidence that the transforming chemicals used in the study did not induce other somatic mutations, it is difficult to rule out the contribution of such mutations in the transformation process. These results nonetheless make a strong case for aneuploidy being a dynamic chromosome mutation event intimately associated with cancer.

Aneuploidy versus somatic gene mutation in cancer

The idea that numeric chromosome imbalance or aneuploidy is a direct cause of cancer was proposed at the turn of the century by Theodore Boveri [33]. However, the hypothesis was largely ignored over the last several decades in favor of the somatic gene mutation hypothesis, mentioned earlier. Evidence accumulating in the literature lately on specific chromosome aneusomies recognized in primary tumors, incidence of aneuploidy in cells undergoing transformation, and aneuploid tumor cells showing a high rate of chromosome instability have led to the rejuvenation of Boveri's hypothesis. The concept has recently been discussed as a "vintage wine in a new bottle" [34•]. The author points out that except for rare cancers caused by dominant retroviral oncogenes, diploidy does not seem to occur in solid tumors, whereas aneuploidy is a rule rather than exception in cancer.

Aneuploidy as an effective mutagenic mechanism driving tumor progression, on the other hand, is being recognized as a viable solution to the paradox that with known mutation rate in non-germline cells ($\sim 10^{-7}$ per gene per cell generation) tumor cell lineages cannot accumulate enough mutant genes during a human lifetime [35]. The concept is gaining significant credibility since genes that potentially affect chromosome segregation were found mutated in human cancer. Some of

these genes have also been shown to have transforming capability in *in vitro* assays. Selected recent publications describing the findings are being discussed below in reference to the mitotic targets potentially involved in inducing chromosome segregation anomalies in cells.

Potential mitotic targets and molecular mechanisms of aneuploidy

Because aneuploidy represents numeric imbalance in chromosomes, it is reasonable to expect that aneuploidy arises due to missegregation of chromosomes during cell division. There are many potential mitotic targets, which could cause unequal segregation of chromosomes (Fig. 1). Recent investigations have identified several genes involved in regulating these mitotic targets and mitotic checkpoint functions, which can be implicated in induction of aneuploidy in tumor cells. This discussion is restricted to those mitotic targets and checkpoint genes whose abnormal functioning has been observed in cancer or has been shown to cause tumorigenic transformation of cells, in recent years. The role of telomeres is discussed elsewhere in this issue. For a more detailed description of the components of mitotic machinery and their possible involvement in causing chromosome segregation abnormalities in tumor cells, readers may refer to a recently published review [36•].

Among the mitotic targets implicated in cancer, centrosome defects have been observed in a wide variety of malignant human tumors. Centrosomes play a central role in organizing the microtubule network in interphase cells and mitotic spindle during cell division. Multipolar mitotic spindles have been observed in human cancers *in situ* and abnormalities in the form of supernumerary

centrosomes, centrosomes of aberrant size and shape as well as aberrant phosphorylation of centrosome proteins have been reported in prostate, colon, brain, and breast tumors [37,38]. In view of the findings that abnormal centrosomes retain the ability to nucleate microtubules *in vitro*, it is conceivable that cells with abnormal centrosomes may missegregate chromosomes producing aneuploid cells. The molecular and genetic bases of abnormal centrosome generation and the precise pathway through which they regulate the chromosome segregation process remain to be elucidated. Recent discovery of a centrosome-associated kinase STK15/BTAK/aurora2, naturally amplified and overexpressed in human cancers, has raised the interesting possibility that aberrant expression of this kinase is critically involved in abnormal centrosome function and unequal chromosome segregation in tumor cells [39,40]. Exogenous expression of the kinase in rodent and human cells was found to correlate with an abnormal number of centrosomes, unequal partitioning of chromosomes during division, and tumorigenic transformation of cells. It is relevant in this context to mention that the *Xenopus* homologue of human STK15/BTAK/aurora2 kinase has recently been shown to phosphorylate a microtubule motor protein XIEg5, the human orthologue of which is known to participate in the centrosome separation during mitosis [41]. Findings on STK15/aurora2 kinase, thus, provide an interesting lead to a possible molecular mechanism of centrosome's role in oncogenesis. Centrosomes have, of late, been implicated in oncogenesis from studies revealing supernumerary centrosomes in *p53*-deficient fibroblasts and overexpression of another centrosome kinase PLK1 being detected in human non-small cell lung cancer [42].

One of the critical events that ensures equal partitioning of the chromosomes during mitosis is the proper and timely separation of sister chromatids that are attached to each other and to the mitotic spindle. Untimely separation of sister chromatids has been suspected as a cause of aneuploidy in human tumors. Cohesion between sister chromatids is established during replication of chromosomes and is retained until the next metaphase/anaphase transition. It has been shown that during metaphase-anaphase transition, the anaphase promoting complex/cyclosome triggers the degradation of a group of proteins called securins that inhibit sister chromatid separation. A vertebrate securin (v-securin) has recently been identified that inhibits sister chromatid separation and is involved in transformation and tumorigenesis. Subsequent analysis revealed that the human securin is identical to the product of the gene called pituitary tumor transforming gene, which is overexpressed in some tumors and exhibits transforming activity in NIH3T3 cells. It is proposed that elevated expression of the v-securin may contribute to generation of malignant tumors due to

Figure 1. Potential mitotic targets causing aneuploidy in oncogenesis

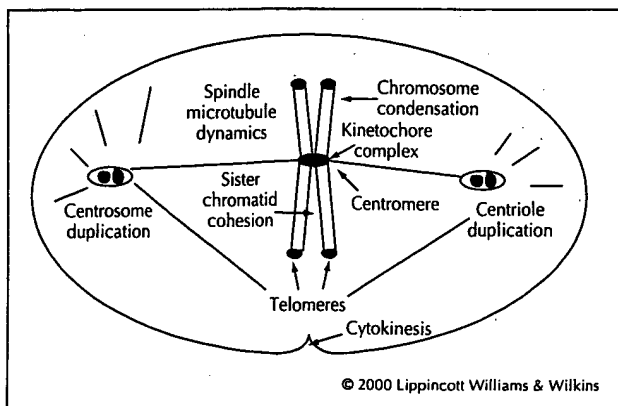


Diagram illustrates that defects in several processes involving chromosomal, spindle microtubule, and centrosomal targets, in addition to abnormal cytokinesis, may cause unequal partitioning of chromosomes during mitosis, leading to aneuploidy. Recently obtained evidence in favor of some of these possibilities is discussed in the text.

chromosome gain or loss produced by errors in chromatid separation [43•].

Normal progression through mitosis during prophase to anaphase transition is monitored at least at two checkpoints. One checkpoint operates during early prophase at G2 to metaphase progression while the second ensures proper segregation of chromosomes during metaphase to anaphase transition. Several mitotic checkpoint genes responding to mitotic spindle defects have been identified in yeast. The metaphase-anaphase transition is delayed following activation of this checkpoint during which kinetochores remain unattached to the spindle. The signal is transmitted through a kinetochore protein complex consisting of Mps1p and several Mad and Bub proteins [44]. It is expected that for unequal chromosome segregation to be perpetuated through cell proliferation cycles giving rise to aneuploidy, checkpoint controls have to be abrogated.

Following this logic, Vogelstein *et al.* [45•] hypothesized that aneuploid tumors would reveal mutation in mitotic spindle checkpoint genes. Subsequent studies by these investigators have proven the validity of this hypothesis and a small fraction of human colorectal cancers have revealed the presence of mutations in either hBub1 or hBubR1 checkpoint genes. It was further revealed that mutant BUB1 could function in a dominant negative manner conferring an abnormal spindle checkpoint when expressed exogenously. Inactivation of spindle checkpoint function in virally induced leukemia has also recently been documented following the finding that hMAD1 checkpoint protein is targeted by the Tax protein of the human T-cell leukemia virus type 1. Abrogation of hMAD1 function leads to multinucleation and aneuploidy [46].

In addition to mitotic spindle checkpoint defects, failed DNA damage checkpoint function in yeast is frequently associated with aberrant chromosome segregation as well. It, therefore, appears intriguing yet relevant that the human *BRCA1* gene, proposed to be involved in DNA damage checkpoint function, when mutated by a targeted deletion of exon 11 led to defective G2/M cell cycle checkpoint function and genetic instability in mouse embryonic fibroblasts [47]. The cells revealed multiple functional centrosomes and unequal chromosome segregation and aneuploidy. Although the molecular basis for these abnormalities is not known at this time, it raises the interesting possibility that such an aneuploidy-driven mechanism may be involved in tumorigenesis in individuals carrying germline mutations of *BRCA1* gene.

Conclusion

Growing evidence from human tumor cytogenetic investigations strongly suggest that aneuploidy is associated with the development of tumor phenotypes. Clinical findings of correlation between aneuploidy and tumorigenesis are supported by studies with *in vitro* grown transformed cell lines. Molecular genetic analyses of tumor cells provide credible evidence that mutations in genes controlling chromosome segregation during mitosis play a critical role in causing chromosome instability leading to aneuploidy in cancer. Further elucidation of molecular and physiologic bases of chromosome instability and aneuploidy induction could lead to the development of new therapeutic approaches for common forms of cancer.

Acknowledgments

The author is thankful to Drs. Bill Brinkley and Pramila Sen for discussions and advice. Help from Ms. Donna Sprabary and Ms. Hongyi Zhou in preparation of this manuscript is gratefully acknowledged. The work in the author's laboratory was supported by grants from the NIH and The University of Texas M.D. Anderson Cancer Center.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- Of special interest
- Of outstanding interest

- 1 Heim S, Mitelman F: Cancer cytogenetics, edn 2. New York: Wiley Liss Inc., 1995.
- 2 Nowell PC: The clonal evolution of tumor cell populations. *Science* 1976, 194:23-28.
- 3 Tomlinson IP, Novelli MR, Bodmer WF: The mutation rate and cancer. *Proc Natl Acad Sci USA* 1996, 93:14800-14803.
- 4 Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. *Nature* 1998, 396:643-649.
- An excellent review on the significance and possible mechanisms of genetic instability in cancer.
- 5 Heppner GH, Miller FR: The cellular basis of tumor progression. *Int Rev Cytol* 1998, 177:1-56.
- 6 Wolman SR: Chromosomal markers: signposts on the road to understanding neoplastic disease. *Diag Cytopath* 1998, 18:18-23.
- 7 Ross JS: DNA ploidy and cell cycle analysis in cancer diagnosis and prognosis. *Oncology* 1996, 10:867-890.
- 8 Magennis DP: Nuclear DNA in histological and cytological specimens: measurement and prognostic significance. *Br J Biomed Sci* 1997, 54:140-148.
- 9 Fletcher JA: Renal and bladder cancers. In: *Human Cytogenetic Cancer Markers*. Edited by Wolman SR, Sell S. Totowa, NJ: Humana Press; 1997:169-202.
- 10 Zhuang Z, Park WS, Pack S, Schmidt L, Vortmeyer AO, Pak E, et al.: Trisomy 7-harboring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nat Genet* 1998, 20:66-69.
- 11 Lindberg JO, Stenling RB, Rutegard JN: DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. *Br J Surg* 1999, 86:947-950.
- 12 Sasaki O, Kido K, Nagahama S: DNA ploidy, Ki-67 and p53 as indicators of lymph node metastasis in early gastric carcinoma. *Anal Quant Cytol Histol* 1999, 21:85-88.
- 13 Abad M, Ciudad J, Rincon MR, Silva I, Paz-Bouza JL, Lopez A, et al.: DNA aneuploidy by flow cytometry is an independent prognostic factor in gastric cancer. *Anal Cell Path* 1998, 16:223-231.
- 14 DeAngelis PM, Clausen OP, Schjølberg A, Stokke T: Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their

- associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999, 80:526-535.
- 15 Malkhosyan S, Yasuda J, Scoto JL, Sekiya T, Yokota J, Perucho M: Molecular karyotype (amplotype) of metastatic colorectal cancer by unbiased arbitrarily primed PCR DNA fingerprinting. *Proc Natl Acad Sci (USA)* 1998, 95:10170-10175.
 - 16 Ai H, Barrera JE, Pan Z, Meyers AD, Varella-Garcia M: Identification of individuals at high risk for head and neck carcinogenesis using chromosome aneuploidy detected by fluorescence in situ hybridization. *Mut Res* 1999, 439:223-232.
 - 17 Barrera JE, Ai H, Pan Z, Meyers AD, Varella-Garcia M: Malignancy detection by molecular cytogenetics in clinically normal mucosa adjacent to head and neck tumors. *Arch Otolaryngol Head Neck Surg* 1998, 124:847-851.
 - 18 Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al.: 17p (p53) allelic loss, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's oesophagus. *Proc Natl Acad Sci USA* 1996, 93:7081-7084.
 - 19 Teodori L, Gohde W, Persiani M, Ferrario F, Tirindelli Danesi D, Scarpignato C, et al.: DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow up study on a cohort of patients. *Cytometry* 1998, 34:257-263.
 - 20 Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, Paulson TG, et al.: Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* 1999, 22:106-109.
 - 21 Sturgis CD, Caraway NP, Johnston DA, Sherman SI, Kidd L, Katz RL: Image analysis of papillary thyroid carcinoma fine needle aspirates: significant association between aneuploidy and death from disease. *Cancer* 1999, 87:155-160.
 - 22 Hemmer S, Wasenius VM, Knuutila S, Joensuu H, Franssila K: Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. *Br J Cancer* 1998, 78:1012-1017.
 - 23 Mendelin J, Grayson M, Wallis T, Visscher DW: Analysis of chromosome aneuploidy in breast carcinoma progression by using fluorescence in situ hybridization. *Lab Inv* 1999, 79:387-393.
 - 24 Bulten J, Poddighe PJ, Robben JC, Gemmink JH, deWilde PC, Hanselaar GAGJM: Interphase cytogenetic analysis of cervical intraepithelial neoplasia. *Am J Pathol* 1998, 152:495-503.
 - 25 Krauter J, Ganser A, Bergmann L, Raghavachar A, Hoelzer D, Lübbert M, et al.: Association between structural and numerical chromosomal aberrations in acute myeloblastic leukemia: a study by RT-PCR and FISH in 447 patients with de novo AML. *Ann Hematol* 1999, 78:265-269.
 - 26 Van Den Neste E, Louviaux I, Michaux JL, Delannoy A, Michaux L, Hagemeijer A, et al.: Myelodysplastic syndrome with monosomy 5 and/or 7 following therapy with 2-chloro-2'-deoxyadenosine. *Br J Hematol* 1999, 105:268-270.
 - 27 Namba M, Mihara K, Fushimi K: immortalization of human cells and its mechanisms. *Crit Rev Oncog* 1996, 7:19-31.
 - 28 Li R, Yerganian G, Duesberg P, Kraemer A, Willer A, Rausch C, Hehlmann R: Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. *Proc Natl Acad Sci USA* 1997, 94:14506-14511.
 - 29 Liu P, Zhang H, McLellan A, Vogel H, Bradley A: Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* 1998, 150:1155-1168.
 - 30 Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC, Altieri DC: Control of apoptosis and mitotic spindle checkpoint survival. *Nature* 1998, 396:580-584.
 - 31 Lengauer C, Kinzler KW, Vogelstein B: Genetic instability in colorectal cancers. *Nature* 1997, 386:623-627.
Demonstrates chromosomal instability in aneuploid colorectal tumor cells.
 - 32 Duesberg P, Rausch C, Rasnick D, Hehlmann R: Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 1998, 95:13692-13697.
Correlates aneuploidy and transformation in *in vitro* grown CHE cells.
 - 33 Boveri T: Zur Frage der Entstehung maligner Tumoren. Jena, Verlag von Gustav Fischer, 1914.
 - 34 Bialy H: Aneuploidy and cancer: vintage wine in a new bottle? *Nat Biotech* 1998, 16:137-138.
Discusses the significance of aneuploidy and gene mutations in cancer.
 - 35 Orr-Weaver TL, Weinberg RA: A checkpoint on the road to cancer. *Nature* 1998, 392:223-224.
 - 36 Pihan GA, Dosey SJ: The mitotic machinery as a source of genetic instability in cancer. *Semin Cancer Biol* 1999, 9:289-302.
Describes various components and regulatory mechanisms of mitotic machinery and possible mechanisms of chromosome missegregation in cancer.
 - 37 Pihan GA, Purohit A, Wallace J, Knecht H, Udda B, Queensberry P, Dosey SJ: Centrosome defects and genetic instability in malignant tumors. *Cancer Res* 1998, 58:3974-3985.
 - 38 Lingle WI, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad Sci USA* 1998, 95:2950-2955.
 - 39 Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al.: Tumor amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998, 20:189-193.
Describes oncogenic property of centrosome associated STK15/aurora2 kinase and its involvement in aneuploidy induction.
 - 40 Bischoff JR, Anderson L, Shu Y, Morsie K, Ng I, Chan CS, et al.: A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998, 17:3052-3065.
Describes oncogenic property of STK15/aurora2 kinase and involvement in colorectal cancers.
 - 41 Giet R, Uzbekov R, Cubizolles F, Le Guellac K, Prigent C: The xenopus laevis aurora related protein kinase pEq2 associates with and phosphorylates the Kinesin related protein X1Eq5. *J Biol Chem* 1999, 274:15005-15013.
 - 42 Zimmerman W, Sparks C, Dosey S: Amorphous no longer: the centrosome comes into focus. *Curr Opin Cell Biol* 1998, 11:122-128.
 - 43 Zou H, McGarry TJ, Bernal T, Kirschner MW: Identification of a vertebrate sister chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* 1999, 285:418-421.
Demonstrates transforming and tumorigenic function of a gene inhibiting sister chromatid separation.
 - 44 Hardwick KG: The spindle checkpoint. *Trends Genet* 1998, 14:1-4.
 - 45 Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JKV, et al.: Mutations of mitotic checkpoint genes in human cancers. *Nature* 1998, 392:300-303.
Describes mitotic checkpoint gene mutations in human colorectal cancers showing chromosome instability.
 - 46 Jin DY, Spencer F, Jeang KT: Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* 1998, 93:81-91.
 - 47 Xu X, Weaver Z, Linke SP, Li C, Gotay J, Wang XW, et al.: Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform deficient cells. *Mol Cell* 1999, 3:389-395.

| | | |
|----|---|---|
| | ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| | REFERENCE | 1 (bases 1 to 437) |
| | AUTHORS | Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Pavello,A., Gish,M., Hawkins M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore B., Morris,M., Parsons,J., Prange,C., Riffin,L., Rolfling,T., Schellendy,K., Soares,M.B., Tan,F., Thirry-Meg,J., Trevaak,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996) |
| | TITLE | 97044478 |
| | JOURNAL | |
| | MEDLINE | |
| | PUBMED | 8889549 |
| | COMMENT | Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: es@watson.wustl.edu This clone is available royalty-free through LINL ; contact the IMAGE Consortium (infoimage.lnl.gov/) for further information. Insert Length: 832 Std Error: 0.00 Seq primer: ml3 -40 forward High quality sequence stop: 375. Location/Qualifiers 1..437 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:3794409" /db_xref="taxon:9606" /clone="IMAGE:245163" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /clone_1lb="Soares fetal liver spleen INFPS" /note="Organ: Liver and Spleen; Vector: pRTD3 (Pharmacia) with a modified polylinker; Site 1: Pac I; site 2: Eco RI 1st strand cDNA was primed with a Pac I - oligo(dT) primer 15' AACTGGAGAAATTAATAAGATCTTTTATTATTTTATTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac and Eco RI sites of the modified pRTD3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo." |
| | FEATURES | |
| | SOURCE | |
| | BASIS COUNT | 78 a 132 c 119 g 106 t 2 others |
| | ORIGIN | |
| | Match | 41.1%; Score 408; DB 14; Length 437; Local Similarity 97.3%; Pred. No. 4,5e-48; |
| | Matches | 425; Conservative 0; Mismatch 11; Indels 1; Gaps 1 |
| Dc | 537 AACAGCGTAGATGGGAAGAATGGGACAGTCCTGCAGAGAGACCATGCCAAGGAG | 596 |
| Dc | 437 AACAGCGTAGATGGGAAGAATGGGACAGTCCTGCAGAGAGACCATGCCAAGGAG | 378 |
| OY | 597 CCTGCCACTTCTCCTCTCCGCGAGCCAGACATGSGACTGTTCGTGGCAAGGCTGCA | 656 |
| Dd | 377 CCGCGCACTTCTCCTCTCCGCGAGCCAGACATGSGACTGTTCGTGGCAAGGCTGCA | 318 |
| OY | 657 AACCAAGCCATATTCAGACAACAGCCCTCAAGTGATGCCAGAGTGTACCAAGAG | 716 |
| Dd | 317 AACCAAGCCATATTCAGACAACAGCCCTCAAGTGATGCCAGAGTGTACCAAGAG | 258 |
| OY | 717 ATGAGAGACTGGAGAGGATCCCTGGAGAGCCCATCTTGCGCGCTCTCACAGAGC | 776 |
| Dd | 257 ATGAGAGACTGGAGAGGATCCCTGGAGAGCCCATCTTGCGCGCTCTCACAGAGC | 198 |
| OY | 777 ACCGCGCTCTGATGAAGAAGAGTGGGAGTTTCAGATAGAGATGGGAGGTC-AG | 835 |
| Dd | 197 ACCGCGCTCTGATGAAGAAGAGTGGGAGTTTCAGATAGAGATGGGAGGTC-AG | 138 |
| OY | 836 AGAAGCAAAGCAAGCAAGCAATGTAATTAACGTCAGAGAGCCAGCAAGCAAGCAAGCA | 895 |

| | | | |
|-------------|-----|--|-----------------------------|
| Dy | 137 | AGGACCCAAAGCAGCGACCATGTAGTAAATGAACCTGCAGAGACCAGACGGCGAAGGA | 78 |
| Oy | 896 | CCTGACGCCATCAAGCGTCGACTTTTGTAATTGGAGTTCATGCAATAATGATGTGTTTTA | 955 |
| Dd | 77 | CTGACAGCCCATCAAGCGTCGACTTTTGTAATTGGAGTTCATGCAATAATGATGTGTTTTA | 18 |
| Oy | 956 | GCTGCTCTTGCACAAA | 972 |
| Dd | 17 | GCTGCTCTTGCACAAA | 1 |
| RESULT 4 | | | |
| H73373 | | | |
| LOCUS | | 480 bp | mRNA linear EST 31-OCT-1995 |
| DEFINITION | | YU48E10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone | |
| ACCESSION | | IMAGE1229387.5, | mRNA sequence. |
| VERSION | | H73373 | |
| KEYWORDS | | H73373.1 GI:1047623 | |
| SOURCE | | EST. | |
| ORGANISM | | Homo sapiens (human) | |
| REFERENCE | | Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 480) Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Pavello,A., Gibb,M., Hawkins ,M., Hultman,M., Kucada,T., Lacy,M., Le,M., Le,N., Mardis,B., Moore, ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,P., Thierry-Mieg,J., Trevaekis,B., Underwood,K., Wohlmann,P., Watson,R., Wilson,R. and Marra,M., Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996) 97044478 | |
| TITLE | | 8889549 | |
| JOURNAL | | Contact: Wilson RK | |
| MEDLINE | | Washington University School of Medicine | |
| PUBMED | | 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 | |
| COMMENT | | Tel: 314 286 1800 | |
| | | Fax: 314 286 1810 | |
| | | Email: est@watson.wustl.edu | |
| | | Insert Size: 1102 | |
| | | High quality sequence stops: 306 | |
| | | Source: IMAGE Consortium, LMU | |
| | | This clone is available royalty-free through LMU ; contact the | |
| | | IMAGE Consortium (infoimage.lmu.gov) for further information. | |
| | | Insert Length: 1102 Std Error: 0.00 | |
| | | Seq primer: ML3RP1 | |
| FEATURES | | High quality sequence stop: 306. | |
| SOURCE | | Location/Qualifiers | |
| | | 1..480 | |
| | | /organism="Homo sapiens" | |
| | | /mol_type="mRNA" | |
| | | /db_xref="GDB:3780483" | |
| | | /db_xref="taxon:9606" | |
| | | /clone="IMAGE:229387" | |
| | | /sex="male" | |
| | | /dev_stage="20 week-post conception fetus" | |
| | | /lab_host="DH10B (ampicillin resistant)" | |
| | | /clone_lib="Soares fetal liver spleen INFLS" | |
| | | /note="Organ: Liver and Spleen; Vector: pTZ19D (Pharmacia) | |
| | | with a modified polylinker; Site_1: Site_2: Eco RI | |
| | | 1st strand cDNA was primed with a Pac I - oligo(3') | |
| | | '5' AACGAGAAGATTAATTAAGATCTTTTATTTTATTTT 3') , | |
| | | double-stranded cDNA was ligated to Eco RI adaptors | |
| | | (Pharmacia), digested with Pac I and cloned into the Pac | |
| | | I and Eco RI sites of the modified pTZ19 vector. Library | |
| | | went through one round of normalization. Library | |
| | | constructed by Bento Soares and M.Patima Bonaldo." | |
| BASE COUNT | | 100 a 149 c 130 g 100 t | 1 others |
| ORIGIN | | | |
| Query Match | | 37.6%; Score 373.2; DB 14; Length 480; | |

Best Local Similarity 92.0%; Pred. No. 3.3e-43;
Matches 439; Conservative 0; Mismatches 29; Indels 9; Gaps 4;

12 GGAACGAGAGGTTCTCACTGCCCCGAGAGAGGCTTACACCCAGAGAGGAGGCTC 71
4 GAAACTAGAGAGGTTCTCACTGCCCCGAGAGAGGCTTACACCCAGAGAGGAGGCTC 63
72 CCGAGGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 131
64 CCGAGGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 123
132 GAAATTAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191
124 GAAATTAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 183
192 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 251
184 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 243
252 GGAACGAGAGGTTCTCACTGCCCCGAGAGAGGCTTACACCCAGAGAGGAGGCTC 311
244 GGAACGAGAGGTTCTCACTGCCCCGAGAGAGGCTTACACCCAGAGAGGAGGCTC 303
312 AACCTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367
304 AACCTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
368 GTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
364 TTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
424 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479
424 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

RESULT 5
H58326 418 bp mRNA linear EST 05-OCT-1995
LOCUS YR2C07.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:206316 5', mRNA sequence.

ACCESSION H58326
VERSION H58326.1 GI:1011158
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 418)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Morris, J., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, B., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Werck EST Project
Unpublished
Contact: Wilson R
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 912
High quality sequence strops: 251
Source: IMAGE Consortium, LML
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.jmi.gov) for further information.
Insert Length: 912 Std Error: 0.00
Seq primer: M13RP1
High quality sequence strop: 251.
Location/Qualifiers
1. 418
/organism="Homo sapiens"
/mol_type="mRNA"

FEATURES
SOURCE

/db_xref="GDB:3775447"
/db_xref="taxon:9606"
/clone="IMAGE:206316"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and spleen. Vector: pTZ19 (Pharmacia) with a modified polylinker. Site 1: Pac I; Site 2: Eco RI; Site 3: Hind III. The cDNA was ligated into the pTZ19 vector (Pharmacia), digested with Pac I and Eco RI and ligated into the pTZ19 vector. Library constructed by Berto Soares and M. Patricia Bonaldo."

Query Match
Best Local Similarity 96.1%; Pred. No. 1.7e-40;
Matches 393; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

279 AAGTGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338
3 AAGTGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62
339 CCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 398
63 CCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
339 GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 458
123 GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182
459 TTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
183 TTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 241
519 GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578
242 GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 301
579 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 637
302 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
638 G-TTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 685
362 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 410

RESULT 6
H74302 582 bp mRNA linear EST 31-OCT-1995
LOCUS YR2C07.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:230132 5', mRNA sequence.

ACCESSION H74302
VERSION H74302.1 GI:1047713
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 582)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chappel, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gibb, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Morris, J., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevaskis, B., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Morris, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

[illegible]

Analysis of Genomic and Proteomic Data Using Advanced Literature Mining

Yanhui Hu, Lisa M. Hines, Haifeng Weng, Dongmei Zuo, Miguel Rivera,
Andrea Richardson, and Joshua LaBaer*

Institute of Proteomics, Harvard Medical School—BCMP, 240 Longwood Avenue, Boston, Massachusetts 02115

Received March 13, 2003

High-throughput technologies, such as proteomic screening and DNA micro-arrays, produce vast amounts of data requiring comprehensive analytical methods to decipher the biologically relevant results. One approach would be to manually search the biomedical literature; however, this would be an arduous task. We developed an automated literature-mining tool, termed MedGene, which comprehensively summarizes and estimates the relative strengths of all human gene–disease relationships in Medline. Using MedGene, we analyzed a novel micro-array expression dataset comparing breast cancer and normal breast tissue in the context of existing knowledge. We found no correlation between the strength of the literature association and the magnitude of the difference in expression level when considering changes as high as 5-fold; however, a significant correlation was observed ($r = 0.41$; $p = 0.05$) among genes showing an expression difference of 10-fold or more. Interestingly, this only held true for estrogen receptor (ER) positive tumors, not ER negative. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative tumors worthy of further examination.

Keywords: bioinformatics • micro-array • text mining • gene-disease association • breast cancer

Introduction

At its current pace, the accumulation of biomedical literature outpaces the ability of most researchers and clinicians to stay abreast of their own immediate fields, let alone cover a broader range of topics. For example, to follow a single disease, e.g., breast cancer, a researcher would have had to scan 130 different journals and read 27 papers per day in 1999.¹ This problem is accentuated with high-throughput technologies such as DNA micro-arrays and proteomics, which require the analysis of large datasets involving thousands of genes, many of which are unfamiliar to a particular researcher. In any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study. The ability to interpret these datasets would be enhanced if they could be compared to a comprehensive summary of what is known about all genes. Thus, there is a need to summarize existing knowledge in a format that allows for the rapid analysis of associations between genes and diseases or other specific biological concepts.

One solution to this problem is to compile structured digital resources, such as the Breast Cancer Gene Database¹ and the Tumor Gene Database.² However, as these resources are hand-curated, the labor-intensive review process becomes a rate-limiting step in the growth of the database. As a result, these

databases have a limited scale and the genes are not selected in a systematic fashion.

An alternative approach is automated text mining: a method which involves automated information extraction by searching documents for text strings and analyzing their frequency and context. This approach has been used successfully in several instances for biological applications. In most cases, it has been applied to extract information about the relationships or interactions that proteins or genes have with one another, in the literature or by functional annotation.^{3–7} Thus far, few publications have applied text-mining to examine the global relationships between genes and diseases. Perez-Iratxeta et al. automatically examined the GO (Gene Ontology) annotation of genes and their predicted chromosomal locations in order to identify genes linked to inherited disorders.⁸

To obtain a more global understanding of disease development, it would be valuable to incorporate information regarding all possible gene-disease relationships, including biochemical, physiological, pharmacological, epidemiological, as well as genetic. This information would enable comprehensive comparisons between large experimental datasets and existing knowledge in the literature. This would accomplish two things. First, it would serve to validate experiments by demonstrating that known responses occur as predicted. Second, it would rapidly highlight which genes are corroborated by the literature and which genes are novel in a given context. We have utilized a computational approach to literature mining to produce a

* To whom correspondence should be addressed: jlabacr@hms.harvard.edu.

comprehensive set of gene-disease relationships. In addition, we have developed a novel approach to assess the strength of each association based on the frequency of citation and co-citation. We applied this tool to help interpret the data from a large micro-array gene expression experiment comparing normal and cancerous breast tissue.

Methods

MedGene Database. MedGene is a relational database, storing disease and gene information from NCBI, text mining results, statistical scores, and hyperlinks to the primary literature. MedGene has a web-based user interface for users to query the database (<http://hipseq.med.harvard.edu/MedGene/>).

Text Mining Algorithms. MeSH files were downloaded from the MeSH web site at NLM (National Library of Medicine) (<http://www.nlm.nih.gov/mesh/meshhome.html>) and human disease categories were selected. LocusLink files were downloaded from the LocusLink web site at NCBI (<http://www.ncbi.nlm.gov/LocusLink/>). Official/preferred gene symbol, official/preferred gene name, and gene alternative symbols and names, all relevant annotations and URLs for each LocusLink record, were collected. Gene search terms were used for literature searching and included all qualified gene names, gene symbols, and gene family terms. Primary gene keys, predominantly qualified gene family terms and gene official/preferred symbols, were used to index Medline records. If the official/preferred gene symbols did not meet the standards to be an index, then qualified gene official/preferred names were used. A local copy of Medline records (up to July, 2002) was pre-selected.

A JAVA module examined the MeSH terms and then indexed each Medline record with the appropriate disease terms. A separate JAVA module was used to examine the titles and abstracts for gene search terms and then to index the gene-related Medline records with the relevant primary gene key(s).

Statistical Methods. For every gene and disease pair, we counted records that were indexed for both gene and disease (double positive hits), for disease only (disease single hits), for gene only (gene single hits), and for neither gene nor disease (double negative hits) to generate a 2×2 contingency table. On the basis of the contingency table framework, we applied different statistical methods to estimate the strength of gene-disease relationships and evaluated the results. These methods included chi-square analysis, Fisher's exact probabilities, relative risk of gene, and relative risk of disease¹⁸ (<http://hipseq.med.harvard.edu/MedGene/>). In addition, we computed the "product of frequency", which is the product of the proportion of disease/gene double hits to disease single hits and the proportion of disease/gene double hits to gene single hits. To obtain a normal distribution, we transformed all the statistical scores using the natural logarithm. We selected the log of the product of frequency (LPF) to validate MedGene and to use for the analysis with the micro-array data. Spearman rank-correlation coefficients were used to assess the linear relationship between LPF and micro-array fold change in expression level.

Global Analysis. Diseases with at least 50 related genes were selected for clustering analysis, and the LPF scores were normalized with total score for each disease. Hierarchical clustering was done with the "Cluster" software and the clustering result was visualized using "TreeView" (<http://rana.lbl.gov/EisenSoftware.htm>).

Breast Tissue Micro-Arrays. Eighty-nine breast cancer samples (79% ER-positive) and 7 normal breast tissue samples were selected from the Harvard Breast SPOR frozen tissue repository and were representative of the spectrum of histological types, grades, and hormone receptor immuno-phenotypes of breast cancer. Biotinylated cRNA, generated from the total RNA extracted from the bulk tumor, was hybridized to Affymetrix U95A oligo-nucleotide micro-arrays. These micro-arrays consist of 12,400 probes, which represent approximately 9000 genes. Raw expression values were obtained using GENE-CHIP software from Affymetrix, and then further analyzed using the DNA-Chip Analyzer (dChip) custom software.

Results

Automated Indexing of Medline Records by Disease and Gene. To study the gene-disease associations in the literature, we first compiled complete lists for human diseases and human genes. To index all Medline records that were relevant to human diseases, the Medical Subject Heading (MeSH) Index of Medline records was utilized. MeSH is a controlled medical vocabulary from the National Library of Medicine and consists of a set of terms or subject headings that are arranged in both an alphabetic and an hierarchical structure. Medline records are reviewed manually and MeSH terms are added to each with software assistance.^{9,10} Twenty-three human disease category headings along with all of their child terms (see the Supporting Information, Supplemental Table 1, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table1.html) were selected from the 2002 MeSH Index creating a list of 4033 human diseases.

No index comparable to the MeSH Index exists for genes, and thus, it was necessary to apply a string search algorithm for gene names or symbols found in Medline text. A complete list of genes, gene names, gene symbols, and frequently used synonyms were collected from the LocusLink database at NCBI,^{11,12} which contains 53,259 independent records keyed by an official gene symbol or name (June 18th, 2002). For the purposes of this study, no distinction was made between genes and their gene products. Authors often use the same name for both, differentiating the two only by the use of italics, if at all. For the intended use of this study, this lack of distinction is unlikely to have a large effect and may in fact be beneficial.

Initial attempts to search the literature using these lists revealed several sources of false positives and false negatives (Table 1). False positives primarily arose when the searched term had other meanings, whereas false negatives arose from syntax discrepancies necessitating the development of filters to reduce these errors. The syntax issues were readily handled by including alternate syntax forms in the search terms. The false positive cases, caused by duplicative and unrelated meanings for the terms, were more difficult to manage. Where possible, case sensitive string mapping reduced inappropriate citations. In many cases, however, this was not sufficient and the terms had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.

For the purposes of data tracking, a primary gene key was selected to represent all synonyms that correspond to each gene. Medline records were indexed with a primary gene key when any synonym for that key was found in the title or abstract. Case-insensitive string mapping was used for all searches except as noted above. No additional weight was

Table 1. Systematic Sources of False Positives and False Negatives in Unfiltered Data*

| source of error | error type | example | filter solution |
|---------------------------------------|----------------|--|------------------------------|
| gene symbol/name is not unique | false positive | MAG—myelin associated glycoprotein MAG—malignancy-associated protein | eliminate this term |
| gene symbol is unrelated abbreviation | false positive | PA—pallid homologue (mouse), pallidin (also abbrev. for Pennsylvania) | eliminate this term |
| gene symbol/name has language meaning | false positive | WAS—Wiskott—Aldrich Syndrome (also the word "was") | case-sensitive string search |
| nonstandard syntax | false negative | BAG-1 instead of BAG1 | add dash term |
| unofficial gene name/symbol | false negative | P53 instead of TP53 | add all gene nicknames |
| nonspecified gene name | false negative | estrogen receptor instead of Estrogen receptor 1 | add family stem term |

* In preliminary studies, Medline was searched for co-occurrence of genes and diseases and the resulting output was evaluated to identify error sources that were amenable to global filters. Each error source is categorized by the type of error it causes: false positives are suggested relationships that are not real and false negatives are real relationships that are underrepresented. The filter solutions used are indicated. Note that in some cases, the filter solution itself introduces error. In general, error rates maximized sensitivity, even at the expense of specificity if needed.

added for multiple occurrences of a term or the co-occurrence of multiple synonyms for the same gene key.

Medline records were searched with all qualified gene identifiers, such as the official/preferred gene symbol, the official/preferred gene name, all gene nicknames and all syntax variants. In situations where there are several members of a gene family or splice variants, some authors prefer to use a shortened gene family name, e.g., estrogen receptor, instead of estrogen receptor 1 (*ESR1*), creating a source of false negatives. For this reason, gene family stem terms were created for all genes that have an alpha or numerical suffix (e.g., *IL2RA*, *TGFB*, *ESR1*, etc.) and then used to search the literature. The family stem terms were handled separately from the specific gene names so that it would be clear when linkages were made to the gene family versus a specific member in that family.

To improve performance and accuracy, some pre-selection was applied to the records that were scanned. First, review articles were eliminated to avoid redundant treatment of citations. Second, non-English journals were removed because the natural language filters were only relevant to English publications. Finally, journals unlikely to contain primary data about gene-disease relationships were also removed (e.g., *Int. J. Health Educ.*, *Bedside Nurse*, and *J. Health Econ.*). Together, these filters reduced the 12 198 221 Medline publications (July 2002) by 37%.

Ranking the Relative Strengths of Gene-Disease Associations. In total, there were 618 708 gene-disease co-citations. In which 16% (8297) of all studied genes had been associated to a disease and 96% (3875) of all diseases had been associated to at least one gene. To rank the relative strengths of gene disease relationships, we tested several different statistical methods and examined the results. With the exception of the relative risk estimates, the methods provided similar results with respect to the rank order of the gene-disease association strengths. However, after comparing the results to other databases and after consulting disease experts, the log of the product of frequency (LPF) was selected for further analysis because it gave the best results overall.

Validation of MedGene. In developing this tool, it was important to minimize the number of missed genes (false negatives) and misclassified genes (false positives). However, in situations when these goals were in conflict, inclusiveness was prioritized. To determine the false negative rate in MedGene, breast cancer was used as a test case because it was associated with more genes than any other human disease and because

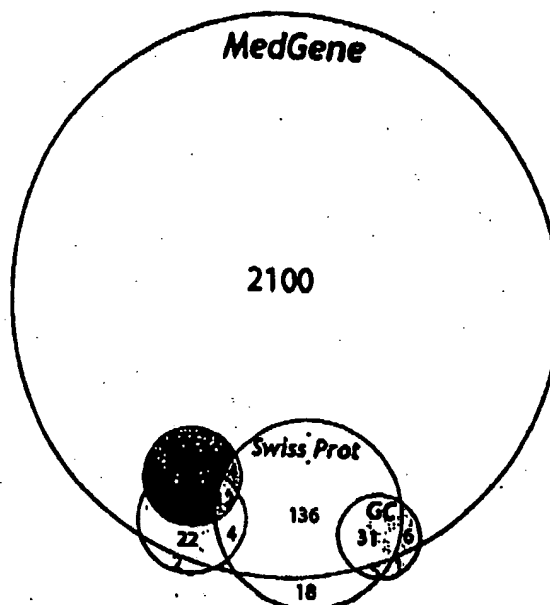


Figure 1. Estimation of the false negative rate by comparison with hand-curated databases. The breast cancer-related genes identified by MedGene were compared with those listed in several other databases including the Tumor Gene Database (TGD),² the Breast Cancer Gene Database (BCG),¹ GeneCards (GC)¹⁷ and Swissprot.¹⁸ Genes were considered false negatives if they were represented in at least one of these other databases and not in MedGene and their link to breast cancer was supported by at least one literature reference. All literature references were verified by manual review to confirm their validity. The number of genes in each database or shared by more than one database is indicated. The false negative rate was calculated by genes missed at MedGene (26)/total number of nonoverlapping genes in other databases (285).

there were several public databases that link genes to breast cancer. We compared the list of breast cancer-related genes from MedGene to these databases, illustrated in Figure 1. Among the 285 distinct breast cancer-related genes that were supported by at least one literature citation in these hand-curated databases, 26 were absent from MedGene, suggesting a false negative rate of approximately 9%. To determine why these were missed, all literature references for these genes (80

papers) were reviewed manually (see the Supporting Information, Supplemental Table 2, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table_2.html). Among these papers, most false negatives were caused by nonstandard gene terms or gene terms eliminated by our specificity filters. Few genes were missed because they were only mentioned in review papers (0.4%) or they appeared only in the body of the manuscript but not the abstract or title (1.1%). Of note, MedGene identified approximately 2000 additional breast cancer-related genes not listed in any other database.

To assess the false positive error rate, two complementary approaches were used: a detailed analysis of one disease and a global examination of 1000 diseases. The detailed approach examined the false positive error rate and its sources, whereas the global approach tested whether the overall results made biomedical sense.

Using the LPF, 1467 genes related to prostate cancer were assembled in rank order. We then retrieved approximately 300 Medline records each for the highest ranked 100 and the lowest ranked 200 genes and manually reviewed the titles and abstracts to determine the verity of the association. Nearly 80% of the highest ranked 100 genes fell into one of the five categories that reflect meaningful gene-disease relationships (see the Supporting Information, Supplemental Table 3, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table_3.html). Among the lowest ranked 200 genes, approximately 70% reflected true relationships. Of the 600 records reviewed, there were only two in which the association between the gene and the disease was described as negative. Both were genes with very low scores. In both cases, the authors did not argue the absence of any relationship, but rather that a particular feature of the gene or protein was not shown to be related to human prostate cancer.^{12,14}

The coincidence of some gene symbols with medical abbreviations, chemical abbreviations and biological abbreviations resulted in most of the false positives (see the Supporting Information, Supplemental Table 4, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table_4.html), emphasizing the importance of the filters that were added in the search algorithm (Table 1). Without the filters, the false positive rate more than doubled, and the false negative rate rose dramatically (data not shown). For example, among the papers about breast cancer, there were only 12 Medline records that referred to *ESR1* and 10 to *ESR2*, whereas almost 2000 papers mentioned estrogen receptor without specifying *ESR1* or *ESR2*; this latter group was detected by the family term filter.

To further validate these results, a global analysis of the gene-disease relationships described by MedGene was performed. For this experiment, it was reasoned that the more closely related the diseases are to one another, the more they will be related to the same gene sets. Thus, if the relationships defined by MedGene accurately reflected the literature, then an unsupervised hierarchical clustering of the gene data should group diseases in a manner consistent with common medical thinking. Conversely, if the clustered diseases do not make sense biologically or medically, it may reflect excessive false positives, false negatives, or inappropriate scoring of the data.

To execute this experiment, the gene sets and the corresponding LPF values for 1000 randomly selected diseases (each with at least 50 gene relationships) were used as a dataset for clustering the diseases. A review of the results showed that the resulting disease clusters were indeed logical based upon common medical knowledge (see the Supporting Information,

Supplemental Figure 1, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_1.html). For example, in one such cluster shown in Figure 2, diabetes and its complications grouped together and were also closely linked to diseases associated with starvation states.

The number of genes associated with a given disease can be estimated by adjusting the MedGene number up by the false negative rate (~9%) and down by the false positive rate (~26% on average). Using this, the average disease has 103.7 ± 45.3 (mean \pm s.d.) genes associated with it, although the range is quite broad with 2359 genes related to breast cancer, 2122 genes related to lung cancer and no genes related to a number of diseases.

Applying MedGene to the Analysis of Large Datasets. Access to a comprehensive summary of the genes linked to human diseases provided an opportunity to analyze data obtained from a high-throughput experiment. We compared the MedGene breast cancer gene list to a gene expression data set generated from a micro-array analysis comparing breast cancer and normal breast tissue samples. Micro-array analysis identified 2286 genes that had greater than a 1-fold difference in mean expression level between breast cancer samples and normal breast samples. Using MedGene, we sorted the 2286 genes into four classes: 555 genes directly linked to breast cancer in the literature by gene term search (first-degree association by gene name); 328 genes directly linked by family term search (first-degree association by family term); 1021 genes linked to breast cancer only through other breast cancer genes (second-degree association); and 505 genes not previously associated with breast cancer. (See the Supporting Information, Supplemental Figure 2, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_2.html.) Among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease. Among the remaining 38 genes, 9 had been related to other cancers, specifically esophageal, colon, uterine, skin, and cervix.

To determine whether the genes highlighted by the micro-array analysis were more likely to have been previously linked to breast cancer in the literature, we created a two-dimensional plot of the fold change of expression level between breast cancer and normal tissue versus the literature score (LPF) (Figure 3A). There was a broad spread of expression changes among the genes directly linked to breast cancer ranging from less than 1-fold change (68%) to over 40-fold (0.3%). Notably, the majority of genes with greater than 10-fold expression changes were linked to breast cancer by first-degree association.

Among all 754 genes directly linked to breast cancer in the literature, there was no correlation between LPF and micro-array fold change ($r = 0.018$, p -value = 0.62). However, when we stratified the analysis based on the magnitude of the fold change, we observed an increasing trend in correlation (Figure 3B) suggesting that genes with a more substantial change in expression level were more likely to have a stronger association in the literature. For genes that had 10-fold change or more in expression level, the correlation increased to 0.41 (p -value = 0.05).

When we evaluated the micro-array data separately for ER positive and ER negative tumors, the trend in correlation between fold change and literature score was highly dependent on estrogen receptor status. Interestingly, there was a similar trend in correlation for ER positive tumors, but no trend in correlation for ER negative tumors.

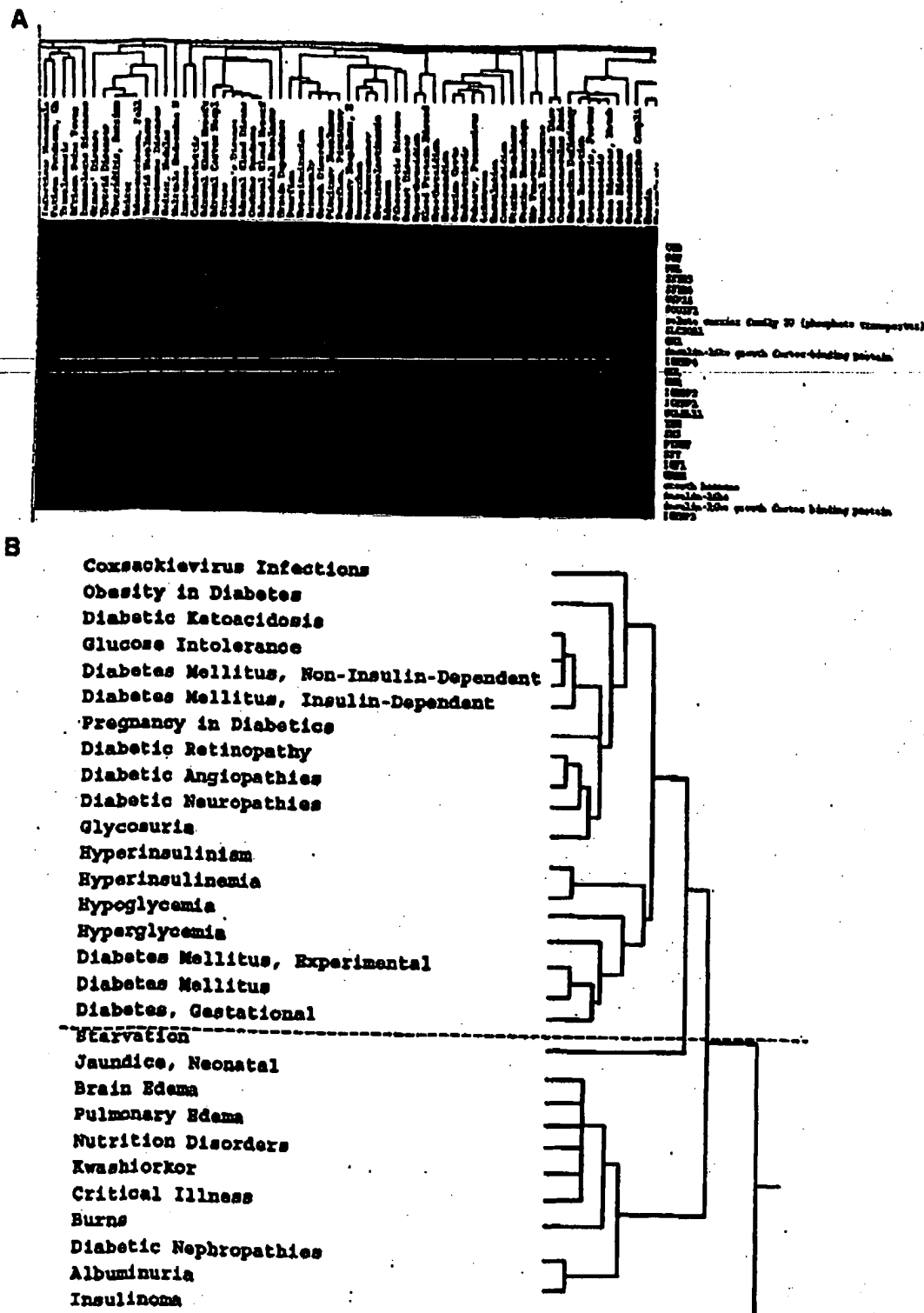


Figure 2. Global validation by clustering analysis. 2(A). The gene sets and the corresponding LPF values for 1000 diseases, each with at least 50 gene relationships, were used in an unsupervised clustering of the diseases based on the gene patterns associated with them. A sample of the data is shown here. 2(B). One of the resulting clusters is shown that corresponds to blood sugar states. Diabetes terms (above the line) and starvation states terms (under the line) clustered together. Within these groups, there is also clustering of diabetic small vessel complications, altered serum chemistries, nutritional disorders, etc. (Supplemental Figure 1: http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_1.html).

Finally, to validate our findings, we computed similar correlations between the breast cancer expression data and LPF scores generated by MedGene for hypertension, a

disease unrelated to breast cancer. As expected, we did not observe an increasing trend in correlation for hypertension.

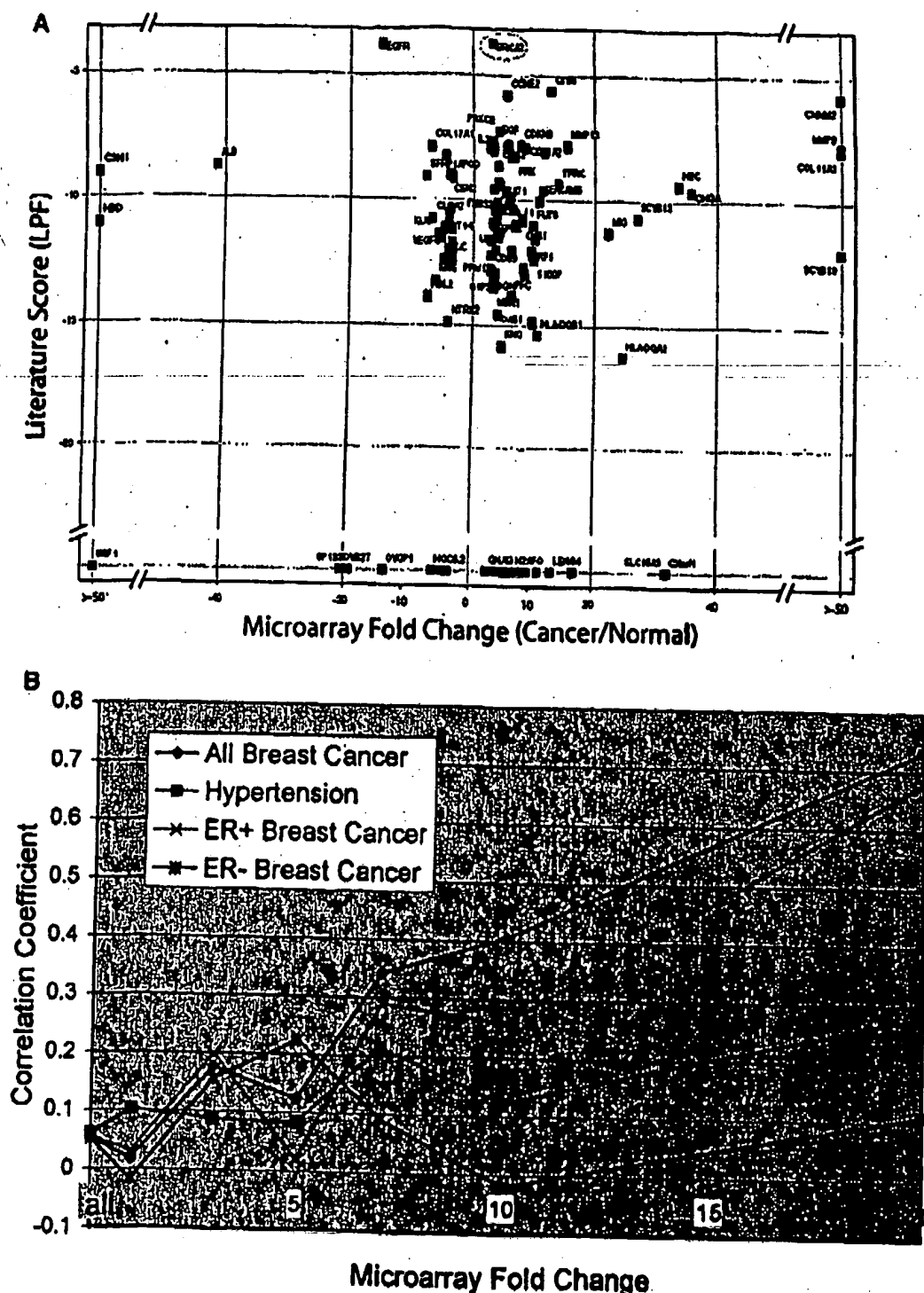


Figure 3. Relationship between literature score and functional data for breast cancer. 3A. The data from an expression analysis of samples for breast tumors and normal breast tissue were analyzed to indicate the fold difference of expression level between breast tumor and normal sample (cutoff ≥ 3 -fold change). The fold changes were plotted against the literature score for the same gene set. Green dots represent first-degree association by gene search, blue dots represent first-degree association by family search and red dots represent no-association. Some well-studied genes, such as BRCA2 (pink circle), are not reflected by a substantial difference in expression changes (shaded area). 3B. The Spearman rank-correlation coefficients between literature score (LPF) and the fold change of expression level between tumor and normal breast samples (y-axis) in relation to the amount of fold change of expression level (x-axis). Gene rank lists were generated for breast cancer (blue) and hypertension (pink). Correlations were also computed between the breast cancer gene LPF scores and fold change expression data among estrogen receptor positive tumors only (light blue) and estrogen receptor negative tumors only (purple).

Table 2. Top 25 Genes Related to Selected Human Diseases*

| breast neoplasms | hypertension | rheumatoid arthritis | bipolar disorder | atherosclerosis |
|-------------------|-------------------------|-------------------------|------------------|---------------------|
| estrogen receptor | <i>REN</i> | <i>RA</i> | <i>ERDA1</i> | apolipoprotein |
| <i>PCR</i> | <i>DBP</i> | <i>TNFRSF10A</i> | <i>SNAP29</i> | <i>APOE</i> |
| <i>ERBB2</i> | <i>LEP</i> | <i>CRP</i> | <i>PFKL</i> | <i>LDLR</i> |
| <i>BRCA1</i> | <i>AGT</i> | <i>AS</i> | <i>DRD2</i> | <i>ELN</i> |
| <i>BRCA2</i> | <i>INS</i> | <i>ESR1</i> | <i>TRH</i> | <i>ARG1</i> |
| <i>EGFR</i> | kallikrein | <i>HLA-DRB1</i> | <i>IMPA2</i> | <i>APOB</i> |
| <i>CYP19</i> | <i>ACE</i> | <i>DR1</i> | <i>HTR3A</i> | <i>APOA1</i> |
| <i>TFF1</i> | endothelin | interleukin | <i>DRD3</i> | <i>MSR1</i> |
| <i>PSEN2</i> | <i>S100A6</i> | <i>TNF</i> | <i>REM</i> | <i>LPL</i> |
| <i>TP53</i> | <i>BDK</i> | <i>IL6</i> | <i>KCNN3</i> | <i>PON1</i> |
| <i>CESS</i> | <i>DIAPH</i> | collagen | <i>DRD4</i> | plasminogen |
| <i>CEACAM5</i> | <i>SAR1</i> | <i>IL1A</i> | <i>HTR2C</i> | activator inhibitor |
| <i>ERBB3</i> | <i>PIH</i> | <i>ACR</i> | <i>RELN</i> | <i>PLG</i> |
| cyclin | <i>CD59</i> | <i>TNFRSF12</i> | <i>DBH</i> | vascular cell |
| <i>COX5A</i> | <i>ALB</i> | <i>IL2</i> | <i>MAOA</i> | adhesion molecule |
| cathepsin | <i>CYP11B2</i> | <i>CHI3L1</i> | <i>COMT</i> | <i>ATOH1</i> |
| <i>ERBB4</i> | <i>MAT2B</i> | <i>IL8</i> | <i>HTR2A</i> | <i>VWF</i> |
| <i>TRAM</i> | angiotensin receptor | interleukin 1 matrix | <i>SYNJ1</i> | <i>INS</i> |
| <i>CCND1</i> | <i>AGTR2</i> | metalloproteinase | <i>INPP1</i> | <i>ARG2</i> |
| <i>EGF</i> | <i>NPPA</i> | interferon | <i>NEDD4L</i> | <i>ABCA1</i> |
| <i>MUC1</i> | <i>LVM</i> | <i>CD68</i> | <i>FRA13C</i> | <i>OLR1</i> |
| Insulin-like | <i>DBH</i> | <i>IL4</i> | transducer of | collagen |
| <i>BCL2</i> | <i>NPY</i> | <i>IL17</i> | <i>ERBB2</i> | <i>MCP</i> |
| mucin | <i>POMC</i> | <i>MMP3</i> | <i>BAIAP3</i> | lipoprotein |
| <i>FGF3</i> | neuropeptide | <i>SIL</i> | <i>ATP1B3</i> | <i>APOA2</i> |
| | | | <i>DRD5</i> | intercellular |
| | | | | adhesion molecule |
| | | | | <i>RAB27A</i> |

* MedGene results for the top 25 genes associated with breast neoplasms, hypertension, rheumatoid arthritis, bipolar disorder, and atherosclerosis, respectively, ranked by LPF scores. The hyperlink to all the papers co-citing the gene and the disease is available at MedGene website (<http://hipseq.med.harvard.edu/MedGene/>).

Discussion

The Human Genome Project heralded a new era in biological research where the emphasis on understanding specific pathways has expanded to global studies of genomic organization and biological systems. High-throughput technologies can provide novel insight into comprehensive biological function but also introduces new challenges. The utility of these technologies is limited to the ability to generate, analyze, and interpret large gene lists. MedGene, a relational database derived by mining the Information in Medline, was created to address this need. MedGene users can query for a rank-ordered list of human gene-disease relationships (Table 2) for one or more diseases. Each entry is hyperlinked to the original papers supporting each association and to other relevant databases.

MedGene is an innovative extension of previous text mining approaches. Perez-Iratxeta et al. used the GO annotation and their chromosomal locations to predict genes that may contribute to inherited disorders.⁸ MedGene takes a broader view and includes all diseases and all possible gene-disease relationships. Furthermore, MedGene utilizes co-citation to indicate a relationship rather than GO annotation, which is limited to the subset of genes that have GO annotation. Our approach is complementary to that taken by Chaussabel and Sher, who used the frequency of co-cited terms to cluster genes into a hierarchy of gene-gene relationships.⁹

A unique aspect of this tool is the ability to assess the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation. This presupposes that most co-citations describe a positive association, often referred to as publication bias¹⁵ and is supported by our observations

that negative associations are rare (Supplemental Table 3: http://hipseq.med.harvard.edu/MedGene/publication/s_Table3.htm). Of course, relationships established by frequency of co-citation do not necessarily represent a true biological link; however, it is strong evidence to support a true relationship.

Another important feature of MedGene is the implementation of software filters that substantially reduced the error rate. We estimate that less than 10% of all associations were missed and at least 70% of even the weakest associations were real. For this study, all of the filters that we applied were general ones, e.g., expanding the list of all gene names to address the different syntax forms used by different journals, eliminating gene names that correspond to common English words, etc. The majority of the remaining search term ambiguities were idiosyncratic and difficult to identify systematically without causing a significant rise in false negatives. Alternative approaches, such as the examination of the nearest neighbor terms, need to be considered to further reduce the false positive rate.

It is not uncommon to see expression changes in micro-array experiments as small as 2-fold reported in the literature. Even when these expression changes are statistically significant, it is not always clear if they are biologically meaningful. When comparing expression levels of disease to normal tissue, one expects an enrichment of known disease-related genes to appear in the altered expression group. MedGene provided a unique opportunity to test this notion in the context of existing knowledge on a novel breast cancer micro-array dataset. For genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a known role in the disease. This

Table 3. Genes with Large Expression Changes in ER- but Not in ER+ Breast Tumors

| gene symbol | fold change (ER+) | fold change (ER-) |
|-------------|-------------------|-------------------|
| KRTHB1 | 1.0 | 610.8 |
| BRS3 | 1.2 | 89.4 |
| DKK1 | 1.2 | 69.8 |
| ZIC1 | 1.9 | 59.6 |
| TLR1 | 1.0 | 38.5 |
| KIAA0680 | 2.8 | 33.2 |
| CDKN3 | 1.0 | 30.6 |
| EBI2 | 4.0 | 27.9 |
| GZMB | 3.8 | 21.9 |
| STK18 | 4.7 | 18.6 |
| GPR49 | 1.0 | 14.6 |
| MYO10 | 1.6 | 14.4 |
| LAD1 | -1.0 | 13.5 |
| POLE2 | 4.2 | 13.0 |
| HMG4 | 4.4 | 12.9 |
| BCL2L11 | -1.2 | 12.3 |
| LRP8 | 2.9 | 12.2 |
| CCNB2 | 1.0 | 11.8 |
| CCNE2 | 4.0 | 11.6 |
| FCB | -4.3 | 11.1 |
| KNSL6 | 2.9 | 10.9 |
| HIF5 | 3.0 | 10.2 |
| SERPINH2 | 4.6 | 10.2 |
| YAP1 | 1.0 | 10.0 |
| LPHB | -1.3 | -10.4 |
| TCEA2 | -1.1 | -10.8 |
| TFF1 | 1.3 | -11.4 |
| COL17A1 | -4.1 | -15.7 |
| POP5 | 1.1 | -16.2 |
| BPAG1 | -4.6 | -22.3 |
| PDZK1 | -1.1 | -36.8 |
| VEGFC | -2.8 | -51.5 |
| MUC6 | -1.4 | -64.9 |
| SERPINA5 | -1.0 | -83.1 |
| MEIS1 | -1.6 | -85.9 |
| CA12 | 2.4 | -150.3 |

Table 3. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative, but not ER positive breast tumors. All of these genes have either never been co-cited with breast cancer or have a weak association except those marked with an *.

reflects the many genes whose role in breast cancer may not involve large changes in expression in sporadic tumors (e.g., *BRCA1* and *BRCA2*) and genes whose modest changes in expression may be unrelated to the disease. Strikingly, among genes with a 10-fold change or more in expression level, there was a strong and significant correlation between expression level and a published role in the disease, providing the first global validation of the micro-array approach to identifying disease-specific genes.

The results derived from MedGene have two implications. First, a careful hunt for corroborating evidence of a role in breast cancer should precede any further study of genes with less than 5-fold expression level changes. Second, any genes with 10-fold changes or more are likely to be related to breast cancer and warrant attention. It is likely that this threshold will change depending on the disease as well as the experiment.

Interestingly, the observed correlation was only found among ER-positive tumors, not ER-negative. This may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently. The MedGene approach identified a set of relatively understudied, yet highly expressed genes in ER-negative tumors that are worthy of further examination (Table 3).

In conclusion, we have developed an automated method of summarizing and organizing the vast biomedical literature. To our knowledge, the resulting database is the most comprehensive and accurate of its kind. By generating a score that reflects the strength of the association, it provides an important tool for the rapid and flexible analysis of large datasets from various high-throughput screening experiments. Furthermore, it can be used for selecting subsets of genes for functional studies, for building disease-specific arrays, for looking at genes common to multiple diseases and various other high-throughput applications. In the future, it will be possible to enhance the utility of the MedGene database by building links between genes and other MeSH terms as well as other biological processes and concepts, such as cell division and responses to small molecules.

Acknowledgment. We would like to thank P. Braun, L. Carraway, J. Pearlberg, and other members of our institute for helpful discussion. Many thanks to the NLM (National Library of Medicine) for licensing of MEDLINE and the annotation effort of adding MeSH indexes for MEDLINE abstracts. This work was funded by grants from the Breast Cancer Research Foundation and an NHLBI PGA Grant (Vol HL66582-02).

Supporting Information Available: Twenty-three human disease category headings along with all of their child terms selected from the 2002 MeSH Index (Supplemental Table 1); analysis of the causes of false negatives in MedGene (Supplemental Table 2); meaningful gene-disease relationships found in MedGene (Supplemental Table 3); causes for incorrect assignment of gene indexes (Supplemental Table 4); a review of the results, showing that the resulting disease clusters were indeed logical (Supplemental Figure 1); and a review of the results showing that among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease (Supplemental Figure 2). This material is available free of charge via the Internet at <http://pubs.acs.org> and at the web sites mentioned in the text.

References

- Basiri, R. A.; Glasser, S. R.; Steffen, D. L.; Wheeler, D. A. *Oncogene* 1999, 18, 7958-7965.
- Steffen, D. L.; Levine, A. E.; Yarus, S.; Basiri, R. A.; Wheeler, D. A. *Bioinformatics* 2000, 16, 639-649.
- Marcotte, E. M.; Xenarios, I.; Eisenberg, D. *Bioinformatics* 2001, 17, 359-363.
- Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. *Bioinformatics* 2001, 17, 155-161.
- Jenssen, T. K.; Laegreid, A.; Komorowski, J.; Hovig, E. *Nat. Genet.* 2001, 28, 21-28.
- Chaussabel, D.; Sher, A. *Genome Biol.* 2002, 3, RESEARCH0055.
- Gibbons, F. D.; Roth, F. P. *Genome Res.* 2002, 12, 1574-1581.
- Perez-Iratxeta, C.; Bork, P.; Andrade, M. A. *Nat. Genet.* 2002, 31, 316-319.
- Funk, M. E.; Reid, C. A. *Bull. Med. Libr. Assoc.* 1983, 71, 176-183.
- Humphrey, S. M.; Miller, N. E. *J. Am. Soc. Inf. Sci.* 1987, 38, 184-196.
- Maglott, D. R.; Katz, K. S.; Scoute, H.; Pruitt, K. D. *Nucleic Acids Res.* 2000, 28, 126-128.
- Pruitt, K. D.; Maglott, D. R. *Nucleic Acids Res.* 2001, 29, 137-140.
- Wadelius, M.; Andersson, A. O.; Johansson, J. E.; Wadelius, C.; Rane, E. *Pharmacogenetics* 1999, 9, 333-340.
- Adam, R. M.; Borer, J. G.; Williams, J.; Eastham, J. A.; Loughlin, K. R.; Freeman, M. R. *Endocrinology* 1999, 140, 5866-5875.
- Montori, V. M.; Smieja, M.; Guyatt, G. H. *Mayo Clin. Proc.* 2000, 75, 1284-1288.
- Denenberg, V. H. *Statistics Experimental Design for Behavioral and Biological Research*; Wiley-Liss: New York, 1976.
- Rebhan, M.; Chalifa-Caspi, V.; Prilusky, J.; Lancet, D. *Trends Genet.* 1997, 13, 183.
- Bairoch, A.; Apweiler, R. *Nucleic Acids Res.* 2000, 28, 45-48, PR0340227.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|-----------|--|
| (51) International Patent Classification ⁵ : C12N 15/11, C12P 21/08 C12Q 1/68, C12N 15/62 C07K 15/28 | A2 | (11) International Publication Number: WO 94/01548 (43) International Publication Date: 20 January 1994 (20.01.94) |
| (21) International Application Number: PCT/GB93/01467 (22) International Filing Date: 13 July 1993 (13.07.93) (30) Priority data: 9214857.6 13 July 1992 (13.07.92) GB (71) Applicant (for all designated States except US): MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): SIBSON, David, Ross [GB/GB]; 37 Grimsdells Lane, Amersham, Buckinghamshire HP6 6HF (GB). GROSS, Jacqueline [GB/GB]; 47 Boxmoor Road, Kenton, Middlesex HA3 8LH (GB). HADFIELD, Kathryn, Mary [GB/GB]; 5 Carlisle Terrace, St Ives, Huntingdon, Cambridgeshire PE17 4PQ (GB). HOWELLS, David [GB/GB]; 77 Puttocks Drive, Welham Green, Hatfield, Hertfordshire AL9 7LW (GB). STARKEY, Michael [GB/GB]; 27 Cressy Close, Abbots Langley, Hertfordshire WO5 0HS (GB). KELLY, Maria [IE/GB]; 24A Oxford Road, Ealing, London W5 3ST (GB). SHAW, Diana [GB/CA]; 342 Glacier Hall, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4 (CA). | | (74) Agent: BIZLEY, Richard, Edward; Hepworth Lawrence Bryer & Bizley, 2nd Floor Gate House South, West Gate, Harlow, Essex CM20 1JN (GB). (81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i> |
| (54) Title: HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE, PLACENTA OR BONE MARROW (57) Abstract <p>This invention provides a nucleic acid fragment encoding a gene product or portion thereof and comprising any one of: (a) a sequence selected from SEQ ID Nos 1 to 1193 from the attached sequence listings; (b) an allelic variation of a sequence as defined in (a); or (c) a sequence complementary to (a) or (b). The invention includes uses of such fragments, and gene products corresponding thereto.</p> | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria | FR | France | MR | Mauritania |
| AU | Australia | GA | Gabon | MW | Malawi |
| BB | Barbados | GB | United Kingdom | NE | Niger |
| BE | Belgium | GN | Guinea | NL | Netherlands |
| BF | Burkina Faso | GR | Greece | NO | Norway |
| BG | Bulgaria | HU | Hungary | NZ | New Zealand |
| BJ | Benin | IE | Ireland | PL | Poland |
| BR | Brazil | IT | Italy | PT | Portugal |
| BY | Belarus | JP | Japan | RO | Romania |
| CA | Canada | KP | Democratic People's Republic of Korea | RU | Russian Federation |
| CF | Central African Republic | KR | Republic of Korea | SD | Sudan |
| CG | Congo | KZ | Kazakhstan | SE | Sweden |
| CH | Switzerland | LI | Liechtenstein | SI | Slovenia |
| CI | Côte d'Ivoire | LK | Sri Lanka | SK | Slovak Republic |
| CM | Cameroon | LU | Luxembourg | SN | Senegal |
| CN | China | LV | Latvia | TD | Chad |
| CS | Czechoslovakia | MC | Monaco | TG | Togo |
| CZ | Czech Republic | MG | Madagascar | UA | Ukraine |
| DE | Germany | ML | Mali | US | United States of America |
| DK | Denmark | MN | Mongolia | UZ | Uzbekistan |
| ES | Spain | | | VN | Viet Nam |
| FI | Finland | | | | |

1
HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN
ADRENAL TISSUE, PLACENTA OR BONE MARROW

This invention relates to new nucleic acid fragments encoding gene products or portions thereof, which fragments are obtainable from human nucleic acid populations, individual members of such populations being present in widely varying amounts.

Situations are increasingly arising in which it is necessary to study complex nucleic acid or polynucleotide populations. For example, it is now widely appreciated that an invaluable resource could be created if the entire sequence of the genomes of organisms such as man were determined and the information available. The magnitude of such a task should not, however, be underestimated. Thus, the human genome may contain as many as 100,000 genes [a very substantial proportion of which may be expressed in the human brain (Sutcliffe, Ann. Rev. Neurosci. 11:157 (1988))]. Only a very small percentage of the stock of human genes has presently been explored, and this largely in a piecemeal and usually specifically targeted fashion.

There has been much public debate about the best means of approaching human genome sequencing. Brenner has argued (CIBA Foundation Symposium 149:6 (1990)) that efforts should be concentrated on cDNAs produced from reverse transcribed mRNAs rather than on genomic DNA. This is primarily because most useful genetic information resides in the fraction of the genome which corresponds to mRNA, and this fraction is a very small part of the total (5% or less). Moreover, techniques for generating cDNAs are also well known. On the other hand, even supposing near perfect recovery of cDNAs corresponding to all expressed mRNAs, some potentially useful information will be lost by the cDNA approach, including sequences responsible for control and regulation of genes. Nonetheless, the cDNA approach at least substantially reduces the inherent inefficiencies resulting from analysis of repeated sequences or non-coding sequences in an approach which depends upon genomic DNA sequencing.

Recently, the results of a rapid method for identifying and characterising new cDNAs has been reported (Adams, M.D. et al., Science 252, 1991, pp 1651-1656). Essentially, a semi-automated sequence reader was used to produce a single read of sequence from one end of each of a number of cDNAs picked at random. It was shown, by comparing the nucleic acid sequences of the cDNAs (or the protein sequences produced by translating the nucleic acid sequence of the cDNAs) to each

other and to known sequences in public databases, that each of the cDNAs picked at random, could be unambiguously classified. The cDNAs could be classified as being either entirely new or as corresponding, to a greater or lesser extent, to a previously known sequence. cDNAs identified in this way were further characterised and found to be useful in a variety of standard applications, including physical mapping. Unfortunately, such a process is insufficient. The longer the process is pursued with any given population of cDNAs the less efficient it becomes and the lower the rate of identification of new clones. In essence, as the number of cDNAs which have already been picked rises, the probability of picking a particular cDNA more than once increases. This difficulty is exacerbated by the wide range of abundancies at which different cDNAs can occur, which abundancies can vary by several orders of magnitude. Thus, whereas some sequences are exceedingly rare, a single cDNA type may comprise as much as 10% of the population of cDNAs produced from a particular tissue (Lewin, B. Gene Expression, Vol. 2: Eukaryotic Chromosomes, 2nd ed., pp. 708-719. New York: Wiley, 1980). The need to avoid missing rarer species in any given population presents a considerable problem.

Various approaches (so-called "normalisation" techniques) have been tried in addressing the problem of increasing the efficiency of examination of a mixed nucleotide population, for example, such a population as is to be examined in human genome sequencing.

Thus, a standard PCR protocol can be used to amplify selectively cDNAs which are present at extremely low levels, if there is information about the sequence of those cDNAs. If not, a primer specific to the desired cDNA cannot be constructed and the desired cDNA cannot be selectively amplified. The standard PCR method is therefore inadequate if it is desired to characterise a number of unknown genes.

A second approach involves hybridization of cDNA to genomic DNA. At saturation, the cDNAs recovered from genomic/cDNA hybrids will be present in the same abundance as the genes encoding them. This will provide a much more homogenous population than the original cDNA library, but does not entirely solve the problem. In order to reach saturation in respect of the very rare sequences, it will be necessary to use huge quantities of cDNA, which need to be allowed to anneal to large amounts of genomic DNA over a considerable periods of time. Furthermore, cDNAs which are homologous to genes which are present in multiple copies in the genome will be over-represented.

A third approach exploits the second order reassociation kinetics of cDNA annealing to itself. After a long period of annealing, the cDNAs which remain single stranded will have nearly the same abundance, and can be recovered by standard PCR (see Patanjali, S.R. et al., PNAS USA 88, 1991, pp. 1943-1947; Ko, M.S.H., NAR 19, no.18, 1991, pp 5705-5711). The methods disclosed in these two publications, however, suffer from notable disadvantages. They are entirely dependent on the stringent physical separation of single stranded and double stranded DNA, require an elevated number of manual manipulations in each reaction, and necessitate protracted hybridisation times (up to 288 hours in the method of Patanjali et al.)

Yet a further approach in "normalising" a nucleotide population is described in co-pending British Patent Application No 91 15407.0, filed 17th July, 1991 by MRC, and involves a PCR process in which a mixture comprising a heterogenous DNA population and appropriate oligonucleotide primers is first formed and the DNA denatured, but before effecting a conventional PCR protocol the conditions are altered to allow the denatured strands of the more common DNA species to reanneal together, whilst avoiding annealing of primers to the DNA strands. By this means, rarer species can subsequently be amplified in preference to the more common species.

This PCR normalisation method in general comprises the steps of:

- (a) preparing a mixture comprising a heterogenous DNA population and oligonucleotide primers suitable for use in a PCR process, in which the DNA is denatured;
- (b) altering the conditions to allow the denatured strands of the more common DNA species to reanneal, while preventing the annealing to the primers to the DNA strands;
- (c) further altering the conditions of the mixture in order to allow the primers to anneal to the remaining single-stranded DNA comprising the rarer DNA species; and
- (d) carrying out an extension synthesis in the mixture produced in step (c).

Advantageously, the method consists of a cyclic application of the above four steps.

It will be appreciated that the conditions may be altered by the alteration of the temperature of the reaction mixture. However, any conditions which affect the hybridisation of complementary DNA strands to one another may be varied to achieve the required result.

5

Because the reannealing efficiency of any given DNA species will depend on the product of its concentration and time, the more abundant the sequence the greater the extent to which it will reanneal in any given time period. Once a DNA species has reached a certain threshold concentration it will no longer be amplified exponentially, as a significant amount will have annealed to the double stranded form before the priming step. Thus, as each individual DNA species is amplified by the process to its threshold concentration, the rate of amplification of that species will start to tail off. Eventually, therefore, all DNA species will be present at the same concentration.

10

15

The length of the reannealing step will determine how much DNA is present at the threshold concentration. Preferably, therefore, the duration of the reannealing step will be determined empirically for each DNA population.

20

In the PCR normalisation process in general, the DNA primers may be adapted to prime selectively a sample of the total DNA population. By using primers which will only prime a sample of the population, only that sample will be amplified and normalised. The total quantity of DNA generated will thereby be reduced, which means that the cycling times can be kept low. This ensures that the method is applicable to complex DNA populations such as cDNA populations. In addition, a first primer can be used which is adapted selectively to prime a sample of the total cDNA population, and a second primer which is a general primer. Advantageously, the general primer is oligo dT (each primed cDNA will then be replicated in its entirety, as the oligo dT primer will anneal to the poly-A tail at the end of the cDNA).

25

30

In co-pending British Patent Application No 92 14873.3, filed by MRC 13th July, 1992, a new process is described which allows the study and identification of the individual members of a mixed or heterogeneous population of nucleotide sequences perhaps of varying abundance. In preferred embodiments of the said process, the starting nucleic acid population is treated by:

40

(a) subjecting the nucleic acid to the action of a reagent,

preferably an endonuclease which has its cleavage and recognition sites separated, which reagent cleaves the nucleic acid so as to produce double stranded cleavage products the individual strands of which overlap at cleaved ends to leave a single strand extending to a known extent;

(b) ligating the cleavage products from (a) with a population of adaptor molecules to generate adapted cleavage products, each of which adaptor molecules has a cleavage product end recognition sequence and the population thereof encompassing a range of adaptor molecules having recognition sequences complementary to a predetermined subset of the sequences of the cleavage-generated extending single strands; and

(c) selecting and separating only those adapted cleavage products resulting from (b) which carry an adaptor of predetermined recognition sequence.

A preferred endonuclease for use in step (a) of the above process is Fok I.

An important feature of this process is the use of adaptor molecules. The adaptors used must have "overhanging" fragment recognition sequences which reflect or are complementary to the extending cleavage-derived sequences which the adaptors are designed to react with. It is also preferred that the adaptors used should end with a 5' hydroxyl group. The avoidance of a 5' phosphate group removes the risk of inappropriate ligation involving the adaptors.

Adaptor molecules may also contain a portion permitting specific sequence selection and separation (as in step (c) of the process) when a sequence is attached to the adaptor. For example, an adaptor can carry biotin, thereby permitting advantage to be taken of the biotin/avidin

reaction in selecting and separating desired adapted molecules. Additionally, adaptors preferably comprise a known and selected sequence such that specifically isolated adapted molecules can be amplified by known techniques (such as PCR) using a primer complementary to the core sequence.

Preferably the adaptors are short double-stranded oligonucleotides which can be joined to the ends of cleavage products. They will have been chemically synthesised so that their sequence can be predetermined

and so that large concentrations can be easily produced. They may also be chemically modified in a way which allows them to be easily purified during the process. As mentioned above, ideally their 5' ends will be unphosphorylated so that once joined to fragments the adapted end of the latter will no longer be able to participate in further ligation reactions.

It is preferred that the adaptor cleavage product end recognition sequences are on the 5' end of the longest oligonucleotide strand making up the preferred adaptor molecules, are at least 3 nucleotides in length and with totally random bases at the single-stranded position(s) two nucleotides in from the 5' end. This then allows selection to be performed both during the joining reaction and during subsequent priming reactions. Then, because the final degree of selection is a result of the product of the degrees of selection achieved at these two stages, maximum selection can be achieved per adaptor/primer available.

Adaptor strand extensions on the 5' end of the longest oligonucleotide also facilitate the use of modified oligonucleotides for separation purposes. Preferably, the short oligonucleotide will be modified at its 5' end. This has the double benefit of requiring just one modified oligonucleotide for all possible single-stranded extensions that are used, and also placing the modification at a position where it cannot interfere with ligation or subsequent priming reactions.

Although only one type of adaptor is required per ligation reaction, it is preferred that adaptors covering all possible reactions in a chosen subset of sequences be present, because then the opportunity for fragments in the chosen subset to ligate to each other is minimised. It is also preferred that the chosen specific adaptor, carrying a predetermined recognition sequence, should not only be different from the other adaptors in its single-stranded extension, but also different in the rest of its sequence since this allows orientation to be introduced which is useful in subsequent steps. It is therefore also preferred that this adaptor has a modified oligonucleotide to facilitate its separation with the cleaving products to which it joins.

The above "adapting" process can be used to generate categories or subsets of sequences by making some of the adaptors specific in some way, and selecting and separating as in step (c). In this way subsets of sequences can be provided depending upon the specific adaptor

chosen, e.g. for use in subsequent nucleotide sequencing. This facilitates, for example, the identification of a large population of sequences by permitting a rational approach to splitting such populations into subsets, each of which subsets can be examined in turn.

In the light of these developments, the present invention now provides a nucleic acid fragment encoding a gene product or portion thereof and comprising any one of:-

- (a) a sequence selected from SEQ ID Nos 1 to 1193;
- (b) an allelic variation of a sequence as defined in (a); or
- (c) a sequence complementary to (a) or (b).

In another aspect, the invention provides a nucleic acid sequence as set out in any one of SEQ ID Nos 1 to 1193, or a complement or allelic variation thereof. Preferred sequences exhibit no more than 90% homology to a human sequence known per se.

In a further aspect, the invention provides a nucleic acid fragment comprising a portion of a sequence as defined above of sufficient size such that a probe of the same size and exhibiting complementarity to said portion can hybridise to said sequence. Preferably, such portions are at least 15 bases in length. It will be appreciated that minor mismatches in the aforesaid "complementarity" are not excluded provided hybridisation can still occur. In general, hybridisation conditions are within the choice of the skilled person, but reference can be made, for example, to the following: Melting temperature of hybrids - Bolton, E. T. and McCarthy, B. J. Proc. Natl. Acad. Sci, 48 p1390 (1962). Effect of formamide on lowering melting temperature - Casey, J. and Davidson, N., Nucleic Acids Res. 4, p1539 (1977). Effect of imperfect homology - Bonner, T. I. et al., J. Mol. Biol. 81, p123 (1973). General - Meinkoth, J. and Wahl, G. Anal. Biochem. 138, p267 (1984). Oligo hybridization and washing - Lathe, R. J. Mol. Biol. 183, P1 (1985).

The present invention also envisages DNA constructs comprising fragments or sequences as referred to above with a control or regulatory sequence.

The invention includes such DNA constructs using a gene system known in the art ligated to a sequence or fragment of the invention so as to enable, upon expression, the provision of a fusion polypeptide. Preferably, an endopeptidase recognition site is provided such that when the sequence or fragment is expressed it is expressed in frame with a known protein with the boundary being a cleavage site for an endopeptidase with a rare cutting site. The known protein can then be affinity purified, and the peptide corresponding to the fragment or sequence in accordance with the invention may be released by the endopeptidase. Alternatively, the whole protein can be used to raise antibodies which can then be screened for those directed at polypeptide corresponding to the fragment or sequence of the invention.

Since the present fragments and sequences can be used to produce, inter alia, corresponding genes, whether by isolating them, by synthesis or otherwise, such use and the resulting DNA fragments comprising genes are further aspects of the invention.

Yet another aspect of the invention is an expression vector comprising a fragment, sequence, gene-comprising DNA fragment, or DNA construct, as above, positioned such that that nucleic acid sequence which encodes the polypeptide corresponding to said fragment, sequence or DNA fragment is in operable reading frame with a control or regulatory sequence.

Other aspects of the invention are host cells incorporating a sequence, or fragment, or gene-comprising DNA fragment, or DNA construct, as above, as a heterologous part of the expressible genetic information of the cell. The production of such modified host cells can be achieved using methods known in the art. Such modified host cells can be used to express corresponding proteins, and these materials lend themselves in turn to the preparation of corresponding monoclonal or polyclonal antibodies using standard techniques.

Also included in this invention are such antibodies. Reference can be made, inter alia, to the following literature: Monoclonal antibodies, Cambell, A. M. Laboratory Techniques in Biochemistry, and Molecular Biology Ed. Burdon, R. H. and van Knippenberg, P. H. vol 13. Elsevier Amsterdam 1984. Goding, J. W., Monoclonal antibodies: Principles and Practice, 2nd Edition, academic Press, London 1986. Kipps, T. J. and Herzenberg, L. A., Handbook of Experimental Immunology : Applications of immunological methods in biomedical sciences, 4th edition Ed. Weir,

D. M. et al., p108 Blackwell scientific Publications, Oxford. Harlow, E and Lane, D. Antibodies, A Laboratory Manual, Cold spring harbor Laboratory, Cold Spring, New York.

5 Expression in an appropriate higher eucaryotic host may be important to ensure correct protein folding and also activity. Expression to avoid copurification of toxic products can sometimes be better performed in organisms approved for human consumption, eg prokaryotic *Bacillus subtilis*, eukaryotic yeast, mammalian cows milk vectors, and other
10 methods known in the art.

The invention also includes novel gene products or portions thereof encoded by a fragment, sequence or gene-comprising DNA fragment of the invention.

15 It will be appreciated that the sequences of the present invention collectively have utility based, inter alia, upon their common origin, and hence they can effectively be considered together rather than as separate entities. It is convenient to represent them as separate
20 sequences, because this is how they were produced, and serves as "punctuation" between the different functional entities which each sequence represents. However, the sequences could just as easily have been presented as a continuous sequence derived by placing them end to end in the order in which they were produced, with a separate
25 indication of where the beginnings and ends of the component sequences are.

In contrast to investigations hitherto, where gene fragments (sequence fragments) could only be identified through some known characteristic
30 [for example: their homology to a fragment which largely encodes amino acids identified by sequencing a previously isolated peptide or is the antisense of that coding sequence; or them having at least partial homology to previously characterised nucleic acids; or them having ability to encode expressed proteins which could later be detected by
35 functional assays of the cells expressing those proteins or by using antibodies which had been previously raised against the proteins to detect their expression, Sambrook J., et al., Molecular Cloning CSH Press 1989], the sequences and fragments described by the present invention are entirely underivable and unpredictable from the prior
40 art, but are nonetheless clearly of great value for various purposes.

Thus, such sequences, by comparing them to sequence databases, can be

used as a means for determining the existence of new members of existing gene families, new human genes when previously only non-human genes were known, and new genes when previously no genes were known (Karlin, S. and Altschul, S. F. Proc. Natl. Acad. Sci. 87 p2264-2268 (1980)). In all cases, this allows the isolation of the corresponding genes and their products, and hence enables the manufacture of molecules of potential biological interest by recombinant means. Screening libraries of known materials or hitherto unexplored source materials for biological efficacy is now an important industrial activity in the search for new therapies and therapeutics. When new sequences have already been found to have counterparts in gene families or in non-human genes then knowledge about biological efficacy may already be apparant. For example, new receptors or receptor agonists/antagonists may exhibit differences to known instances of these molecules, and such differences could make them more suitable as therapeutics by, for example, exhibiting binding characteristics which are more in keeping with avoidance of toxicity. Reference can be made, for example, to polymorphic dopamine receptors and the implications for mental health (Iversen, L. Nature 358, p109 (1992), and Van Tol, H. M. M. et al., Nature 358, p149-152 (1992)). Where absolutely required, realisation of full length cDNAs for expression can be achieved by using the sequences to screen (by hybridisation) suitable cDNA libraries containing full length clones (D'Alession, J. M., et al., Focus (Gibco B.R.L) 9 pl (1987)). Alternatively, the sequences can be used to design primers suitable for obtaining the missing sequences by PCR or other amplification methods (Frohman, M. A., Dush, M. K. and Martin, G. R., Proc. Natl. Acad. Sci. 85 p8998-9002 (1988)).

Appropriate use of the sequence fragments in antisense or triple helix (Griffin et al., Science 245 p967-971 (1989)) applications will be useful for identifying manipulable targets related to disease. For example, viruses have been inhibited by antisense RNA to their mRNAs (Chang, L-J., and Stoltzfuz, C. M. J. Virol. p921-974 (1987)). A similar effect could be achieved by targetting the expression of cellular proteins which are essential for growth or maintenance of the virus.

Partial or full length cDNAs have great utility once expressed. The manner of expression can be selected by one skilled in the art to suit the intended application. Expression of full length cDNAs is typically required for biological activity. Prokaryotic, and lower or higher eucaryotic hosts may be selected as the host for expression and higher

eucaryotes may be preferred to ensure correct modifications, for example, glycosylation in vivo, when this proves to be important. Expression can be ensured by situating the cDNA appropriately to signals for expression (Amann, E. and Brosius, J. Gene 40 p183 1985), Shimuzu, Y et al., Gene 65, p141 (1982), Straus, D. and Gilbert, W. Proc. Natl. Acad. Sci. 22, p2014 (1925)). Such signals may include a promoter for transcription, which may itself be regulatable.

The proteins thus-expressed can be screened for activities of therapeutic or commercial value. It may be that the proteins have to be first isolated for this purpose or can be assayed in situ. It may be desirable that some means of stabilising the expressed protein is employed. This can be achieved, for example, (and as indicated earlier) by expressing in frame as part of a fusion polypeptide (Smith, D. B., et al., Proc. Natl. Acad. Sci. 83 p8073 (1986)).

Useful antibodies can be raised against the expressed proteins. It is commonly not an absolute requirement that full length proteins are produced, although this may influence the quality of the antibodies produced. Peptides as short as 8 or 9 amino-acids in length can be used as antigens (Germain R., N. Nature 353 pp605-607 (1991), Rudensky, A., Y., et al., Nature 353 p622-627 (1991)). Immunogenic peptides could simply be synthesised using the amino-acid sequence translated from a sequence or fragment of this invention. It is desirable, although not absolutely required, that some means of producing purified antibodies is adopted. When fusion polypeptides are used to raise antibodies, an affinity matrix specific for the generic part of the protein allows the fusion polypeptide to be immobilised (Smith, D. B., et al., Proc. Natl. Acad. Sci. 83 p8073 (1986)). The immobilised polypeptide can then be used to affinity purify the antibodies. Antibodies to both the generic part of the fusion polypeptide and the part of interest are produced. When these need to be discriminated between, a different affinity column can be used to remove only those antibodies specific for the generic part of the polypeptide. Alternatively, and as mentioned earlier, it can be arranged that the boundary between the two separate protein components of the fusion polypeptide has the recognition sequence for an endopeptidase with a rare cutting site. The peptide of interest can then be released from the affinity purified polypeptide by the action of the endopeptidase (Nagai, K., and Thorgeren, H., C. Methods Enzymol. 153 p461-481 (1987)). Another alternative is raise monoclonal antibodies against the purified protein.

The antibodies can be used for localising in situ, or quantifying in samples through, for example, ELISA or RIA assays, peptides against which they were raised. These uses are particularly beneficial when the results of the assays can be correlated to a disease condition, eg cancer. For example tumour markers may be found and used to target therapeutic agents. The antibodies can also be used to detect or monitor markers of undifferentiated growth, infection, cardiovascular or immune disease or a therapeutic response. When the antibodies recognise cell surface proteins they can be used in isolation or in combination to isolate particular populations of cells. These in turn can be used to isolate yet more cDNAs which will be enriched for yet more of such surface markers for the population, which, if similarly screened, will permit yet further subdivision of the population. Ultimately, panels of antibodies which can describe particular disease states will accrue. Such antibodies could be tailored for forensic applications as well as diagnostic purposes and disease monitoring.

The sequences or fragments can also be used for genetic analysis and mapping, for example, to diagnose the likelihood that a given individual is predisposed towards a given genetic disease. In the event of a sequence co-locating, genetically, with a disease gene, it can be used for the derivation of new disease therapies bases upon precise genetic knowledge. Such therapies can include, for example, the techniques of so-called "gene therapy" (Dusty Miller, A. Nature 357 p455-460 (1992)).

Antibodies can be produced against the protein of a genetic disease with sufficient discriminating power to discriminate between diseased and non-diseased states (Caskey, T. Genome Sequencing Conference, Hilton Head, S. Carolina (1991)). This would be useful for reducing the dependence of such tests on nucleic acid-based screens. Such antibodies also have the advantage of allowing detection of faulty expression of the protein, for example levels of expression which may be important for development of the disease in slow onset conditions.

Also very important is that not all cDNAs are likely to be found by conventional means, whereas the present sequences are, in one sense, "comprehensive". The use of the class of cDNAs which corresponds of necessity to truncated clones increases the chances that part of a cDNA will be cloned free of any sequences that could otherwise compromise it from being cloned. Sequence obtained can then be used to generate PCR primers from which the remainder can be obtained without having to

clone.

This invention will now be further described and illustrated by means of the following Examples.

5

All oligonucleotides used in these Examples were synthesised Trityl on using an ABI 380B DNA Synthesizer according to the manufacturers instructions. Purification was by reverse phase HPLC (see, for example, Becker, C., R., et al., J. Chromatography 326, p293-299 (1985)).

10

Example 1

Human brain and adrenal tissues were obtained from a mixture of 12 to 15 week menstrual age fetuses and then snap frozen in liquid nitrogen before storing in bijou bottles in a -80°C freezer. The two types of tissue were used separately, directly from the freezer, to prepare cDNA from which restriction fragments were generated for sorting into subsets. 1g portions of each of the separate tissues were homogenised, using an Ultra-Turrax T25 Disperser (Janke and Kunkel, IKA-Labortechnik), on ice in the presence of 4M guanidinium isothiocyanate to solubilise macromolecules. RNA was isolated from each homogenate by using centrifugation to sediment it through caesium trifluoroacetate. This was performed using the Pharmacia kit according to the manufacturer's instructions, except that centrifugation was performed for 36 hours and the RNA obtained was finally desalted and concentrated by performing two ethanol precipitations in succession with two 70% ethanol washes after each precipitation. In each case, polyA⁺ (mRNA) was isolated from 200 to 400 µg of the total RNA by binding it to magnetic oligo-dT coated beads (Dynal). Solution containing unbound material was removed from the beads, which were washed, and then mRNA eluted directly for use. mRNA isolation was performed in accordance with the manufacturer's instructions. Yields of RNA from the beads were between 1 and 3% of the total RNA. 2 to 4 µg of the eluted RNA were used for cDNA synthesis. cDNA synthesis was performed according to the method of Gubler, U and Hoffman, (B. J. Gene 25 p263 (1983)) using a Pharmacia kit according to the manufacturer's instructions. OligodT was used to prime the first strand cDNA synthesis reaction. The cDNA was purified by extracting twice with phenol/chloroform and then low molecular weight solutes including nucleic acids below ca. 300 bases were removed by passing the cDNA reaction mixture through a Pharmacia S400 spun column used according to the manufacturer's

15

20

25

30

35

40

instructions. Running buffer for the column comprised 10 mM TrisHCl, 1 mM EDTA, 50 mM NaCl @ pH 7.5.

5 The column eluate was adjusted to 10 mM Mg^{2+} and then the purified cDNA was restricted by the action of 1 unit per 10 μ l of the endonuclease Fok I at 37°C for 1 hour, so that it would be able to accept adaptors to enable fragment sorting.

10 The cDNA fragments were purified by two successive phenol/chloroform extractions followed by passing them through S400 spun columns as described above.

The adaptors used were oligonucleotides 5' N₃N₃N₃TCCTTCTCCTGCGACAGACA (SEQ ID: 1194) with the complementary strand 5' TGTCTGTGCGAGGAGAAGGA (SEQ ID: 1195) and 5' AAN₃N₃TCTCGGACAGTGCTCCGAGAAC (SEQ ID: 1196) or 5' TTN₃N₃TCTCGGACAGTGCTCCGAGAAC (SEQ ID: 1197) each with the complementary 5' biotinylated strand GTTCTCGGAGCACTGTCCGAGA (SEQ ID: 1198). These were added to 25% of the eluted material by incubating together 200 pmoles of the mixture of double-stranded adaptors in the elution buffer to which had been added $MgCl_2$ to 10mM, ATP to 10mM and 0.025 units/ μ l of T4 DNA ligase. The oligonucleotide 5' biotinylated GTTCTCGGAGCACTGTCCGAGA, (SEQ ID: 1198) and whichever of the complementary oligonucleotides with which it was used, each comprised 1/32 of the molar proportion total adaptors. The final reaction volume was 90 μ l which was heated to 65°C for 3 minutes and then cooled to room temperature before the ligase was added. Ligation was performed for 16 hours at 12°C.

30 Two successive phenol/chloroform extractions were performed to remove the ligase. The final aqueous phase was passed through an S400 spun column (Pharmacia) as described above except that the column was used with 10 mM Tris pH 8.3/50 mM NaCl.

35 The column eluate was adjusted to 25mM Mg^{2+} , 0.5mM dNTPs in a final volume of 200 μ l. The mixture was placed in a thermocycler (Techne MW2) and heated to 78°C for 5 minutes. At this point 10 units of cloned Taq DNA polymerase (AmpliTaq, Perkin Elmer) were added. This was followed by an incubation at 72°C for 10 minutes to fill in the unligated strand of the adaptor. After the second incubation 200 μ l of streptavidin coated magnetic beads (Dynal) prepared according to the manufacturers instructions were added to bind cDNA ligated to that of the oligonucleotides which was complementary to the 5'

GTTCTCGGAGCACTGTCCGAGA (SEQ ID: 1198) biotinylated adaptor. Bead binding was allowed to proceed at 28°C for 30 minutes with mixing every 10 minutes.

- 5 Un-biotinylated cDNAs were washed from the beads with 400µl each of 2M NaCl twice, fresh 0.15 mM NaOH four times at 28°C for 5 minutes each, water twice and finally a buffer comprising 20 mM Tris pH 8.3, 50 mM NaCl, and 25mM Mg²⁺. The beads were then resuspended in 240 µl of the final buffer including additionally 0.5 mM dNTPs and divided into 4x60
10 µl.

- Four of the 60 µl aliquots, two from each tissue, were processed further specifically to prime and copy a subset of the immobilised, adapted fragments. 2 pmoles of the primer 5' CTGTCTGTCCGAGGAGAAGGAA
15 (SEQ ID: 1201) were added to each of two aliquots, one from each tissue. 2 pmoles of the primer 5' CTGTCTGTCCGAGGAGAAGGAG (SEQ ID: 1202) were added to each of the other two aliquots. 2.5 units of Taq DNA polymerase were added to each reaction and 16 cycles of alternate denaturation at 95°C for 30 seconds, annealing at 63°C for 2 minutes
20 and polymerisation at 72°C for 3 minutes was performed to accumulate the selected single-strands in solution.

- On completion of the DNA synthesis reactions a further 30 µl of resuspended beads were added to each reaction to remove the
25 biotinylated fragments. The reaction was incubated at 28°C for 30 minutes mixing every 10 minutes to ensure that the biotinylated strands were bead bound. Each aqueous phase containing the newly synthesised strands was then removed and extracted with phenol/chloroform twice to remove the enzyme before being further purified by passing through an
30 S400 spun column equilibrated with 10 mM Tris pH 8.3/50 mM NaCl as described above.

- Rounds of PCR amplification of subsets of the selected fragments were performed by using the original primer in each case, together with one
35 of the primers 5' GTTCTCGGAGCACTGTCCGAGAG (SEQ ID: 1199) or 5' GTTCTCGGAGCACTGTCCGAGAC (SEQ ID: 1200). This simultaneously rendered the fragments double-stranded and increased the amounts of available material. It was not known how many cycles of amplification would be required at this stage, since each primer pair would be expected to
40 behave differently. It was therefore necessary directly to determine a suitable number empirically by using standard agarose gel electrophoresis to examine the reaction products after a given number

of cycles. In some cases to avoid the accumulation of non-specific products, it was necessary to perform an initial 5 cycles of amplification with both of the primers present at 2 pmoles each. All reactions were performed using 8 μ l or 12.5 μ l whichever was the larger but not exceeding 12 μ l of the column effluent above. Reaction conditions were adjusted to 20 mM Tris pH 8.3, 50 mM NaCl, 25mM Mg^{2+} , 0.5mM dNTPs and 2.5 units of Taq DNA polymerase in a final volume of 40 μ l. Apart from when an initial amplification with 2 pmoles of each primer was performed, 20 pmoles of each primer were used. Cycles of amplification were performed at 95°C for 30 seconds, 65°C for 1 minute and 72°C for 3 minutes.

For the purposes of cloning, selected cDNA was amplified as described immediately above, except that the reaction was not monitored. Instead, the number of cycles which had previously been shown to just give rise to all observable products plus another 4 cycles were performed. In addition, an extra 72°C for 10 minutes incubation was performed after the last cycle.

The products of the reaction were then prepared for directional cloning. Water was added to adjust the final reaction volume to 60 μ l. Enzyme was removed by two successive phenol/chloroform extractions. The final aqueous mixture was passed through an S400 column as described above, except that it had been equilibrated with 10 mM Tris HCl pH 7.5, 50mM NaCl.

For directional cloning, advantage was taken of the different known sequences introduced at each end of the selected cDNAs by the adaptors in a modification of the method of Aslandis, C. and de Jong, P. J. (Nucl. Acids Res. 18, p6156 (1990)). Different cohesive ends were produced on each end by using the exonuclease activity of T4 DNA polymerase to resect from the 3' end, to the first T in each case. To 75 μ l or 75 % of the column eluate, whichever was least, were added 9.5 μ l of 100mM TrisHCl pH7.4, 100 mM $MgCl_2$, and 9.5 μ l of 0.5 mM dTTP. 16 units of T4 DNA polymerase were added and the reaction incubated in a water bath at 37°C for 30 minutes. The enzyme was removed by extracting with phenol/chloroform, twice successively. The salt of the final aqueous phase was adjusted by passing it through an S300 column (Pharmacia) equilibrated with 10 mM TrisHCl pH 7.4, 1 mM EDTA as described above.

The E.coli plasmid cloning vector pBluescript KS+ (Alting-Meese, M. A.

and Short J. M., Nucl. Acids Res. 17 p9494) was prepared for accepting the resected cDNA by restriction cleavage at the BamHI and HindIII sites and then adaptor the resultant cohesive ends using the specific adaptors produced by the oligonucleotide 5' AGCTCGGCTCGAGTCTG (SEQ ID: 1203) with its partially complementary oligonucleotide 5' GCGACAGACAGCAGACTCGAGCCG (SEQ ID: 1204) and the oligonucleotide 5' GATCCGGCTCGAGT (SEQ ID: 1205) with its partially complementary oligonucleotide 5' CCGAGAACACTCGAGCCG (SEQ ID: 1206). Preparation of the vector and adaptor were performed according to standard procedures. Insertion of the cDNA was performed between the BamHI and HindIII restriction sites. Recombinant vectors were transformed into the host XL1-Blue (Bullock, W. O. et al Biotechniques 5 p376-378 (1987)) by the method of Hannahan, D. J. (Mol. Biol. 166 p577-580 (1983)). Suitable standard controls for the ligations and transformations were also included.

Post transformation procedures were as described in "Molecular Cloning", 2nd Edition (Sambrook J., Fritsch, E. F., and Maniatis, T. CSH Press (1989)). Colonies were produced by plating onto X-gal/IPTG L-agar plates containing 50µg/ml ampicillin and 10µg/ml tetracyclin. Clear colonies were picked, each into a separate well of a microtitre plate, containing 100µl of L-broth and 50µg/ml ampicillin. Growth was allowed to occur for 16 hours at 37°C. 100µl of 50% or 30% glycerol was added to plates which were archived at -20°C or -80°C, respectively.

Bacteria corresponding to those archived were used for preparing templates for sequencing by the dideoxy method (Sanger, F. Milklen, S. and Coulson, A. R. Proc. Natl. Acad. Sci. 74 p5463-5467 (1977)). Bacteria for this purpose were either grown on L-agar plates containing 50µg/ml of ampicillin, prepared at the same time as they had been grown in liquid culture, or after plating out from the archive. Alternatively, fresh liquid cultures were inoculated from the archive.

In all cases, cDNA inserts were amplified for sequencing by PCR (Saiki, R. K. et al Science 239 p487-491 (1988)). PCR was either performed using bacteria directly added to the reaction, by a toothpick, or PCR was performed using 1/50th of the plasmid isolated by preparative methods (Holmes, D. S. and Quigley, M. Anal. Biochem. 114 p193 (1981)) from the bacteria in the liquid cultures or from the plates.

20 pmoles of each of the PCR primers 5' biotinylated GTAAAACGACGGCCAGT

(SEQ ID: 1207) and 5' CGAGGTCGACGGTATCG (SEQ ID: 1208) were used in 40µl reactions containing 2.5mM Mg²⁺, 50 mM KCl, Tris-HCl pH 8.3 and 0.25 units of Amplitaq (Cetus). Reactions were performed at 95°C, for 1 minute, followed by 35 cycles at 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 40 seconds. After the cycles, a final incubation at 72°C for 5 minutes was performed.

After PCR, standard agarose gel electrophoresis was used to determine which reactions had been successful. The biotinylated strands of successful reactions were then recovered for single-stranded sequencing by binding them to streptavidin coated beads (Dynal) and then washing, all according to the manufacturers instructions, except that the washing steps were either performed manually or performed automatically in the 96 well microtitre plate format using a Biomek robotic workstation attached to a side-arm loader (Beckman).

Dideoxy chain termination sequencing reactions were performed using the immobilised, biotinylated strands as templates and 2 pmoles of the oligonucleotide 5' CGAGGTCGACGGTATCG (SEQ ID: 1209) as primer. Reactions were performed using fluorescently-labelled terminators (Du Pont) or a fluorescein-labelled primer (Pharmacia) according to the manufacturers instructions. Reactions were analysed using automated DNA sequencers. A Genesis 2000 was used for the "Du Pont" reactions and an A.L.F. for the "Pharmacia" reactions. Bases were assigned for the Genesis 2000 reads using the manufacturers Base Caller software. Files of called bases were then transferred to a SUN Network from an Apple Macintosh computer which had been used for base calling. Raw data from the A.L.F. reads was directly transferred to a SUN network where bases were called using the public domain "trace editor software" (TED). In both cases, files of called bases were entered into a Sybase™ database. Entering data entailed automatically removing vector and adaptor or linker sequences, but not editing ambiguous bases. After removal of the unwanted bases, files were automatically compared to other sequences in the cDNA database and the latest versions of the publically available databases, GENBANK and SWISSPROT. Searches were performed with the "basic local alignment search tool" (BLAST) (Karlin, S. and Altschul, S. F. Proc. Natl. Acad. Sci. 87 p2264-2268 (1990)).

Sequences SEQ ID Nos 1 to 610, given hereinafter, were obtained by the above procedure.

Example 2

A second method of preparing cDNA libraries for obtaining gene fragments of the invention took advantage of the PCR normalisation process described above. Standard procedures were used to prepare mRNA from RNA that had been isolated by standard caesium chloride bouyant density gradient methods from a full term human placenta. The oligonucleotide LNotdt, sequence 5' TACGTTGACAAGCTTGAATTCGGGCCCGC(T)_n, (SEQ ID: 1210) was used at 1 μ M with AMV reverse transcriptase, to prime first strand cDNA synthesis under standard conditions from 0.5 μ g of the placental mRNA. Temperatures above 65°C were used to inactivate the reverse transcriptase and then the volume of the reaction made up to 100 μ l with water.

PCRs were then performed in reactions containing 1 μ l of the diluted cDNA, 10 mM Tris-HCl pH 8.3, 40 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin, 200 μ M dNTPs, 10 uCi α^{32} P dCTP, 1 μ M each of the primers 11AD1, sequence 5' GCC(TA)(GC)CGCCGA (SEQ ID: 1211), and LNotdT and Taq DNA polymerase. An initial denaturation period of 95°C for 90 seconds was followed either by 35 cycles of standard PCR, comprising 95°C for 30 seconds, 45°C for 30 seconds and 72°C for 30 seconds or alternatively 3 cycles of the standard PCR already described followed by 27 cycles of Cot PCR during which an additional step of 72°C for 16 minutes was placed between all of the 95°C and 45°C steps of the standard PCR. The standard PCR was followed by a single 72°C for 3 minutes step while the Cot PCR was followed by one standard PCR cycle except that the 72°C incubation was performed for 3 minutes.

Products of the PCR reaction were end repaired by adding 5 units of T4 DNA polymerase to the reaction and then incubating at 37°C for 10 minutes. Enzymes were removed by phenol extraction. The cDNA was precipitated by 70% ethanol, dried and then resuspended in NotI buffer. 20 units of NotI were used to digest the cDNA under standard conditions. cDNA was again phenol extracted and ethanol precipitated. 10% of the purified NotI cut DNA were ligated to the vector pBluescript lting-Meese, M. A. and Short J. M. Nucl. Acid Res. 17 p9494 which had been prepared as standard to receive this DNA by restricting with the enzymes NotI and EcoRV. Transformation and processing of clear colonies was performed as described above except that the host E.coli strain DH5a was used in place of XL-1 Blue.

Preparation of clones for sequencing, sequencing and sequence analysis of cDNAs in clones thus-produced were performed as described in Example 1.

Sequences SEQ ID Nos 611 to 772, given hereinafter, were obtained by the above procedure.

Example 3

5

cdNA libraries corresponding to adult brain cortex (Clontech Laboratories, Inc., Cat No. HL10036) and adult bone marrow (Clontech Laboratories, Inc., Cat No. HL10586) prepared in lambda gt11 phage were transfected into E.coli Y1090 and plated out for colour selection of recombinant plaques ("Molecular Cloning", 2nd Edition Sambrook J., Fritsch, E. F., and Maniatis, T. CSH Press (1989)). 192 lambda Zap clones, corresponding to rhabdomyosarcoma cDNAs and a gift from C. Cooper, ICR, Sutton, were similarly plated except that the host XL-1 Blue was used.

15

Clear plaques from each library were resuspended in 5 μ l of Tris-HCl pH 8, 1 mM EDTA. 2 μ l of the resultant phage suspensions were added directly to PCRs for the purpose of amplifying the cDNA inserts for sequencing. PCR was performed as described in Example 1, except that the oligonucleotides used as primers for the lambda gt11 clones were 5' GGTGGCGACGACTCCTGGAGCCCG (SEQ ID: 1212) and 5' TTGACACCAGACCAACTGGTAATG (SEQ ID: 1213). Whichever of the oligonucleotides was to be used to prime the strand which would serve as the sequencing template was used in biotinylated form.

25

Preparation of clones for sequencing, sequencing and sequence analysis of cDNAs in clones thus-produced was performed as described in Example 1, except that 2 pmoles of the primers that were unbiotinylated in the PCR were used as sequencing primers.

30

Sequences SEQ ID Nos 773 to 1193, given hereinafter, were obtained by the above procedure.

35 The following are the SEQUENCE LISTINGS which comprise sequences SEQ ID Nos 1 to 1213 referred to hereinbefore. Certain of these sequences are preferred, and are listed as such after the main SEQUENCE LISTINGS.

CLAIMS:

1. A nucleic acid fragment encoding a gene product or portion thereof and comprising any one of:-
 - (a) a sequence selected from SEQ ID Nos 1 to 1193;
 - (b) an allelic variation of a sequence as defined in (a); or
 - (c) a sequence complementary to (a) or (b).
2. A nucleic acid sequence as set out in any one of SEQ ID Nos 1 to 1193, or a complement or allelic variation thereof.
3. A sequence as claimed in claim 2 and which exhibits no more than 90% homology to a human sequence known per se.
4. A nucleic acid fragment comprising a portion of a sequence as defined in claim 2 or claim 3 of sufficient size such that a probe of the same size and exhibiting complementarity to said portion can hybridize to said sequence as defined in claim 2 or claim 3.
5. A fragment as claimed in claim 4, wherein said portion is at least 15 bases in length.
6. A fragment as claimed in any one of claims 1, 4 or 5 and encoding at least a portion of a biologically active polypeptide.
7. A nucleic acid sequence as claimed in claim 2 or claim 3 and encoding at least a portion of a biologically active polypeptide.
8. A DNA construct comprising a fragment as defined in any one of claims 1, 4, 5 or 6 or a sequence as defined in any one of claims 2, 3 or 7, together with a control or regulatory sequence.
9. A construct as claimed in claim 8 which encodes a fusion protein comprising a known protein and the polypeptide encoded by said fragment or sequence.
10. A construct as claimed in claim 9, wherein the fusion protein encoded is a cleavable fusion protein having an endopeptidase recognition site positioned between codons corresponding to said known

614

protein and said fragment or sequence.

11. The use of a fragment as defined in any one of claims 1, 4, 5 or 6 or a sequence as defined in any one of claims 2, 3 or 7 to produce a
5 gene.

12. A DNA fragment comprising a gene obtainable by the use defined in claim 11.

10 13. An expression vector comprising a fragment as defined in any one of claims 1, 3, 5 or 6, a sequence as defined in any one of claims 2, 3 or 7, a DNA construct as defined in any one of claims 8 to 10, or a DNA fragment as claimed in claim 12, positioned such that that nucleic acid sequence which encodes the polypeptide corresponding to said
15 fragment, sequence or DNA fragment is in operable reading frame with a control or regulatory sequence.

14. A vector as claimed in claim 13, wherein said vector control or regulatory sequence comprises a regulatable promoter.
20

15. Host cells which incorporate as a heterologous part of their expressible genetic information a fragment as defined in any one of claims 1, 3, 5 or 6, a sequence as defined in any one of claims 2, 3 or 7, or a DNA fragment as defined in claim 12.
25

16. A process for the production of a polypeptide comprising cultivating host cells as defined in claim 15.

17. An antibody directed against a polypeptide obtainable by the performance of a process as defined in claim 16.
30

18. An antibody as claimed in claim 17 and which is monoclonal.

19. A novel gene product or portion thereof encoded by a fragment as defined in any one of claims 1, 3, 5 or 6, or encoded by a sequence as defined in any one of claims 2, 3 or 7, or encoded by the gene
35 comprised in a DNA fragment as defined in claim 12.

40

<first sequence: ss.H74302 (length = 582)
<second sequence: ss.DNA57836 (length = 992)

<503 matches in an overlap of 562: 89.50 percent similarity
<gaps in first sequence: 0, gaps in second sequence: 13 (20 bases)
<score: 1385 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:36:56 2005]: 1 sequence found

H74302 yu56c11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
IMAGE:230132 5', mRNA sequence. 582 bp,
mRNA, linear, EST 31-OCT-1995

ACCESSION H74302
VERSION H74302.1 GI:1047713
KEYWORDS EST; 5_prime.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 582)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1114
High quality sequence stops: 308
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1114 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 308.

FEATURES
source Location/Qualifiers
 1..582
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3781228"
 /db_xref="taxon:9606"
 /clone="IMAGE:230132"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal liver spleen 1NFLS"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

```

              10      20      30      40      50
ss.H74302      CAGGAAGTAGGAGGTTCTCACTGCCCCAGCAGAXGGCCCTACACCCACCGA
                *****
ss.DNA57836    GGCACGAGCCAGGAAGTCTCACTGCCCCAGCAGA-GGCCCTACACCCACCGA
              10      20      30      40      50

              60      70      80      90     100     110
ss.H74302      GGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTTTCTCAA
                *****
ss.DNA57836    GGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGC-TTCTCAA
            60      70      80      90     100     110

              120     130     140     150     160     170
ss.H74302      GGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTT
                *****
ss.DNA57836    GGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTT
            120     130     140     150     160     170

              180     190     200     210     220     230
ss.H74302      CCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCAC
                *****
ss.DNA57836    CCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCAC
            180     190     200     210     220     230

              240     250     260     270     280     290
ss.H74302      CTATTCCTCTGTGGAACCAAGAACATCAAGGTTGCCAAGAAGGTGGTGAAGACCCACGA
                *****
ss.DNA57836    CTATTCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGA
            240     250     260     270     280     290

              300     310     320     330     340     350
ss.H74302      GCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTXCAGTCCAGACCTGGCTCACCTAAT
                *****
ss.DNA57836    GCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCT-GCTCACCTACT
            300     310     320     330     340     350

              360     370     380     390     400     410
ss.H74302      TTATGGCCGGGCGTCCTCCACCTXAGGTGCCCATGTGGGACAGTTXCCAGGGTTACAGAT
                *
                *****
ss.DNA57836    T--CTGCCGGGCGTCCTCCACCTCAGGTGCCCATGT-GGACAGT--GCCAGGCTACAGAT
            360     370     380     390     400     410

              420     430     440     450     460     470
ss.H74302      GCATTGGGGAGGTTGTGGTTCCAAGCCAGTGTXTGAGGTTGCGGGGXCAATTTXAATTTT
                ***
                *****
ss.DNA57836    GCACTGGG--AGCTGTGG-TCCAAGCCAGTGTCTGAGCT--GCGGGCCAACCTTCACTCT
            420     430     440     450     460

              480     490     500     510     520     530
ss.H74302      GXAGGACAGAGGGGXAAAGGXCCAGGGTTXGGAGXTGATTTTGCCAGGXTTTTtagggxag
                *
                *****
ss.DNA57836    GCAGGACAGAGGGGCAGGCCCCAGGGT--GGAGATGA-TCTGCCAGGCGTCCTCGGGCAG
            470     480     490     500     510     520

              540     550     560     570     580
ss.H74302      TCCAATTATTXATCAAAAAGTTXATTTXGGGAAGXTTTGGXAAGGTTXAAT
                ***
                *****
ss.DNA57836    CCCACCTA-TCACCAACAGCCTGA--TCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAG
            530     540     550     560     570     580

ss.DNA57836    ACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCTGCCGAGCCAGACATCGGACTGGTT
            590     600     610     620     630     640

```

ss.DNA57836 CTGGTGCCAGGCTGCAAACAAGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCC
650 660 670 680 690 700

ss.DNA57836 AGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTT
710 720 730 740 750 760

ss.DNA57836 GCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCAGGATAGG
770 780 790 800 810 820

ss.DNA57836 GAATGGGGAGGTGAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCC
830 840 850 860 870 880

ss.DNA57836 AAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAA
890 900 910 920 930 940

ss.DNA57836 AATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA
950 960 970 980 990

<first sequence: ss.H74303 (length = 414)
<second sequence: ss.DNA57836 (length = 992)

<196 matches in an overlap of 414: 47.34 percent similarity
<gaps in first sequence: 8 (20 bases), gaps in second sequence: 3 (6 bases)
<score: 474 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:35:36 2005]: 1 sequence found

H74303 yu56c11.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
IMAGE:230132 3', mRNA sequence. 414 bp,
mRNA, linear, EST 31-OCT-1995

ACCESSION H74303
VERSION H74303.1 GI:1047714
KEYWORDS EST; 3_prime.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 414)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
PUBMED 8889549
COMMENT Contact: wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1114
High quality sequence stops: 313
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1114 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 313.

FEATURES
source Location/Qualifiers
 1..414
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3781228"
 /db_xref="taxon:9606"
 /clone="IMAGE:230132"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal liver spleen 1NFLS"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

```

ss.DNA57836  GGCACGAGCCAGGAAGTCTCTACTGCCCCAGCAGAGGCCCTACACCCACCGAG
                10      20      30      40      50      60

ss.DNA57836  GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG
                70      80      90     100     110     120

ss.H74303
                                                    GTG

ss.DNA57836  CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTC
                130     140     150     160     170     180

ss.H74303  GCAAGAGCAGCTAAAACACACTCATTTCATGAAGTCCAAATACGAACAGTGCACGCTG
                10      20      30      40      50      60
*** ** ** * * * * *

ss.DNA57836  CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT
                190     200     210     220     230     240

ss.H74303  ATGGC--CTGCAGTCCTCTGCGTGCTTGGCTCTCTGGACGGTTTCATTCTACATGGCTGC
                70      80      90     100     110     120
** * ** * * * * *

ss.DNA57836  ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC
                250     260     270     280     290     300

ss.H74303  TGCTTTGCGTC-CTCTGACCTCCCCATTC---CCTATCCTGAACCCCCCAACTCCTCT
                130     140     150     160     170
* * * * *

ss.DNA57836  CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT
                310     320     330     340     350     360

ss.H74303  TC-----ACTCAGACGGCGGGTGCTCCTGTAGAGCG-GCAAGGCAAGGATGGGGCT--C
                180     190     200     210     220
* * * * *

ss.DNA57836  GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG
                370     380     390     400     410     420

ss.H74303  TCCAGGGGACCCTGCCAGTCCCTCATCTTCTGGTCACCACCTGGGGGCACCACTGTGAG-
                230     240     250     260     270     280
* * * * *

ss.DNA57836  AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACCTTCACTCTGCAGGACAGAGGGG
                430     440     450     460     470     480

ss.H74303  --GGCGCTGTGCTGGACATTGGCGTTGTTTGCAGCCTGGCACCAGAACAGTCCGATGTC
                290     300     310     320     330     340
*** * * * *

ss.DNA57836  CAGGCCCCAGGGTGGAGA-TGATCTGCCAGGCGTCTCGGGCAGCCACCTATCACCAAC
                490     500     510     520     530

ss.H74303  TGGCTCGGCAGGAAGGAGAAGTTGGCAGG-CTGCCTGTGGATXGGTCTCTGCTGCAGGTG
                350     360     370     380     390     400
* * * * *

ss.DNA57836  AGCCTGATCGGGAAGGA----TGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCA
                540     550     560     570     580     590

ss.H74303  GACCTGCCCAT
                410
* * * * *

ss.DNA57836  G-CCTGCCAACTTCTCCTTCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTG
                600     610     620     630     640     650

ss.DNA57836  CAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCAGGTGGTGACCAGA
                660     670     680     690     700     710

```

ss.DNA57836 AGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCCTTGCCGCTCTACAGGA
720 730 740 750 760 770

ss.DNA57836 GCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCA
780 790 800 810 820 830

ss.DNA57836 GAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGG
840 850 860 870 880 890

ss.DNA57836 ACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTT
900 910 920 930 940 950

ss.DNA57836 AGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAAA
960 970 980 990

<first sequence: ss.H73373 (length = 480)
<second sequence: ss.DNA57836 (length = 992)

<456 matches in an overlap of 471: 96.82 percent similarity
<gaps in first sequence: 0, gaps in second sequence: 7 (9 bases)
<score: 1303 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:35:00 2005]: 1 sequence found

H73373 yu48f10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
IMAGE:229387 5', mRNA sequence. 480 bp,
mRNA, linear, EST 31-OCT-1995

ACCESSION H73373

VERSION H73373.1 GI:1047623

KEYWORDS EST; 5_prime.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 480)

AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
PUBMED 8889549

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1102
High quality sequence stops: 306
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1102 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 306.

FEATURES Location/Qualifiers

source 1..480
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3780483"
 /db_xref="taxon:9606"
 /clone="IMAGE:229387"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal liver spleen 1NFLS"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

```

          10      20      30      40      50
ss.H73373      CAGGAAACTAGGAGGTTCTCACTGCCCCGAGCAGAGGCCCTACACCCACCGAG
          *  *  *****
ss.DNA57836      GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCCGAGCAGAGGCCCTACACCCACCGAG
          10      20      30      40      50      60

          60      70      80      90      100     110
ss.H73373      GCATGGGGCTCCCTGGGCTGTTCTGCTTGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG
          *****
ss.DNA57836      GCATGGGGCTCCCTGGGCTGTTCTGCTTGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG
          70      80      90      100     110     120

          120     130     140     150     160     170
ss.H73373      CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC
          *****
ss.DNA57836      CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC
          130     140     150     160     170     180

          180     190     200     210     220     230
ss.H73373      CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT
          *****
ss.DNA57836      CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT
          190     200     210     220     230     240

          240     250     260     270     280     290
ss.H73373      ATTCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC
          *****
ss.DNA57836      ATTCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC
          250     260     270     280     290     300

          300     310     320     330     340     350
ss.H73373      CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTTCACCTXATT
          *****
ss.DNA57836      CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGC-TCACCTACTT-
          310     320     330     340     350

          360     370     380     390     400     410
ss.H73373      CTTGCCGGGCGTTCCTCCACCTCAGGTGCCCATGTGGGACAGTGCCAGGGTTACAGATGG
          *****
ss.DNA57836      -CTGCCGGGCG-TCCTCCACCTCAGGTGCCCATGT-GGACAGTGCCA-GGCTACAGATGC
          360     370     380     390     400     410

          420     430     440     450     460     470
ss.H73373      CATTGGGGAGCTGTGGTCCAAGCCAGTGTTTTGAGGTGCGGGCCAATTTTCATTTTGCAGG
          *  *****
ss.DNA57836      ACT--GGGAGCTGTGGTCCAAGCCAGTG-TCTGAGCTGCGGGCCAACCTTCACTCTGCAGG
          420     430     440     450     460     470

          480
ss.H73373      ACAGAGGG
          *****
ss.DNA57836      ACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTA
          480     490     500     510     520     530

ss.DNA57836      TCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACA
          540     550     560     570     580     590

ss.DNA57836      GGCAGCCTGCCAACTTCTCCTTCTGCGGAGCCAGACATCGGACTGGTTCTGGTGCCAGG
          600     610     620     630     640     650

ss.DNA57836      CTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACC
          660     670     680     690     700     710

```


ss.DNA57836 AGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACA
720 730 740 750 760 770

ss.DNA57836 GGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCAGGATAGGGAATGGGGAGG
780 790 800 810 820 830

ss.DNA57836 TCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAG
840 850 860 870 880 890

ss.DNA57836 AGGACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAAAATGAGTGTGT
900 910 920 930 940 950

ss.DNA57836 TTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAAA
960 970 980 990

<first sequence: ss.H58326 (length = 418)
<second sequence: ss.DNA57836 (length = 992)

<400 matches in an overlap of 415: 96.39 percent similarity
<gaps in first sequence: 1 (1 base), gaps in second sequence: 4 (4 bases)
<score: 1155 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:32:19 2005]: 1 sequence found

H58326 yr25c07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
IMAGE:206316 5', mRNA sequence. 418 bp,
mRNA, linear, EST 05-OCT-1995
ACCESSION H58326
VERSION H58326.1 GI:1011158
KEYWORDS EST; WASHU_est; 5_prime; merck_est.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 418)
AUTHORS Hillier,L., Wilson,R., etal
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT High quality stops: 251; stops: 251; insert: 912.
FEATURES Location/Qualifiers
source 1..418
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3775447"
/db_xref="taxon:9606"
/clone="IMAGE:206316"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen 1NFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

| | | | | | | |
|-------------|----------------|---------------|---------------|--------------|-------------|-----|
| ss.DNA57836 | GGCACGAGCCAGGA | ACTAGGAGGTTCT | CACTGCCCCGAGC | AGAGGCCCTAC | ACCCACCGAG | |
| | 10 | 20 | 30 | 40 | 50 | 60 |
| ss.DNA57836 | GCATGGGGCTCCCT | GGGCTGTTCTGCT | TGGCCGTGCTGG | CTGCCAGCAGCT | TCTCCAAGG | |
| | 70 | 80 | 90 | 100 | 110 | 120 |
| ss.DNA57836 | CACGGGAGGAAGAA | ATTACCCCTGTGG | TCTCCATTGCCT | ACAAAGTCCTG | GGAAGTTTTCC | |
| | 130 | 140 | 150 | 160 | 170 | 180 |
| ss.DNA57836 | CCAAAGGCCGCTGG | GTGCTCATAACCT | GCTGTGCACCCC | AGCCACCCACCG | CCCCATCACCT | |
| | 190 | 200 | 210 | 220 | 230 | 240 |
| ss.H58326 | | | | 10 | 20 | |
| | | | | GAAGGTGGGTGA | AGACCCACGAG | |
| | | | | ***** | ***** | |
| ss.DNA57836 | ATTCCTCTGTGGA | ACCAAGAACATCA | AGGTGGCCAAGA | AGGT-GGTGAAG | ACCCACGAG | |
| | 250 | 260 | 270 | 280 | 290 | |

| | | | | | | |
|-------------|---|-----|-----|-----|-----|-----|
| ss.H58326 | 30 | 40 | 50 | 60 | 70 | 80 |
| | CCGGCCTCCTTCAACCTCAACGTCAACTCAAGTCCAGTCCAGACCTGCTCACCTACTTC | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 300 | 310 | 320 | 330 | 340 | 350 |
| ss.H58326 | 90 | 100 | 110 | 120 | 130 | 140 |
| | TGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGG | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 360 | 370 | 380 | 390 | 400 | 410 |
| ss.H58326 | 150 | 160 | 170 | 180 | 190 | 200 |
| | GAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGG | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 420 | 430 | 440 | 450 | 460 | 470 |
| ss.H58326 | 210 | 220 | 230 | 240 | 250 | 260 |
| | GCAGG-CCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAAC | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 480 | 490 | 500 | 510 | 520 | 530 |
| ss.H58326 | 270 | 280 | 290 | 300 | 310 | 320 |
| | AGCCTGATCGGGAAGGATGGGCAGGTCCACCTXCAGCAGAGACCATGCCACAGGAGGCCT | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 540 | 550 | 560 | 570 | 580 | 590 |
| ss.H58326 | 330 | 340 | 350 | 360 | 370 | 380 |
| | GCCAATTGTCCTTCCTTGCCGAGCCAGACATCGGACTGGTTTTTGGTGCCAGGGTGCAA | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 600 | 610 | 620 | 630 | 640 | 650 |
| ss.H58326 | 390 | 400 | 410 | | | |
| | ACAACGTCAATGTCCAGXACAGGGXCCTTTAGAXTT | | | | | |
| ss.DNA57836 | ***** | | | | | |
| | 660 | 670 | 680 | 690 | 700 | 710 |
| ss.DNA57836 | ACAACGCCAATGTCCAGCACAGCG-CCCTCACAGTGGTGCCCCAGGTGGTGACCAGAAG | | | | | |
| | 720 | 730 | 740 | 750 | 760 | 770 |
| ss.DNA57836 | ATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGC | | | | | |
| | 780 | 790 | 800 | 810 | 820 | 830 |
| ss.DNA57836 | ACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCAGGATAGGGAATGGGGAGGTCAGA | | | | | |
| | 840 | 850 | 860 | 870 | 880 | 890 |
| ss.DNA57836 | GGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGAC | | | | | |
| | 900 | 910 | 920 | 930 | 940 | 950 |
| ss.DNA57836 | TGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAG | | | | | |
| | 960 | 970 | 980 | 990 | | |
| ss.DNA57836 | CTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA | | | | | |

<first sequence: ss.R02548 (length = 379)
<second sequence: ss.DNA57836 (length = 992)

<205 matches in an overlap of 379: 54.09 percent similarity
<gaps in first sequence: 10 (107 bases), gaps in second sequence: 5 (14 bases)
<score: 443 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:31:10 2005]: 1 sequence found

R02548 ye80a07.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
IMAGE:124020 3', mRNA sequence. 379 bp,
mRNA, linear, EST 31-MAR-1995
ACCESSION R02548
VERSION R02548.1 GI:752284
KEYWORDS EST; WASHU_est; 3_prime; merck_est.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 379)
AUTHORS Hillier,L., Wilson,R., etal
TITLE The washU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT High quality stops: 309; stops: 309; insert: 846.
FEATURES Location/Qualifiers
 source 1..379
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:476565"
 /db_xref="taxon:9606"
 /clone="IMAGE:124020"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal liver spleen 1NFLS"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

| | | | | | | |
|-------------|----------------|---------------|----------------|----------------|---------------|------|
| ss.DNA57836 | GGCACGAGCCAGGA | ACTAGGAGGTTCT | CACTGCCCGAGCAG | AGGCCCTACACCC | ACCGAG | |
| | 10 | 20 | 30 | 40 | 50 | 60 |
| ss.DNA57836 | GCATGGGGCTCCCT | GGGCTGTTCTGCT | TGGCCGTGCTGGCT | GCCAGCAGCTTCT | CCAAGG | |
| | 70 | 80 | 90 | 100 | 110 | 120 |
| ss.DNA57836 | CACGGGAGGAAGAA | ATTACCCCTGTGG | TCTCCATTGCCTAC | AAAGTCCTGGAAG | TTTTTC | |
| | 130 | 140 | 150 | 160 | 170 | 180 |
| ss.DNA57836 | CCAAAGGCCGCTGG | GTGCTATAACCTG | TGTGTCACCCCAG | CCACCACCGCCCA | TACACCT | |
| | 190 | 200 | 210 | 220 | 230 | 240 |
| ss.R02548 | | | | | 10 | |
| | | | | | TTTTTGTGGCAAG | AGC |
| | | | | | * * * * | **** |
| ss.DNA57836 | ATTCCTCTGTGGA | ACCAAGAACATCA | AGGTGGCCAAGAAG | GGTGGTGAAGACCC | ACGAGC | |
| | 250 | 260 | 270 | 280 | 290 | 300 |

| | | | | | | |
|-------------|---|-----|-----|-----|-----|-----|
| ss.R02548 | 20 | 30 | 40 | 50 | 60 | 70 |
| | AG--CTAAAACACACTCA--TTTTCATGAACCTCAAATACGAACAGTGCACGCTGAT-G | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT | | | | | |
| | 310 | 320 | 330 | 340 | 350 | 360 |
| ss.R02548 | 80 | 90 | 100 | 110 | 120 | 130 |
| | GCCTGCAGTCCTCTGCCGTGCTTGGCTCTCTGGACAGT-TCATTCTACATGGCTGCTGCT | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACA-----GATGCA | | | | | |
| | 370 | 380 | 390 | 400 | | 410 |
| ss.R02548 | 140 | 150 | 160 | 170 | 180 | |
| | TTGCGTCCTCT-GACCTCCCCATTCCCTATCCT--GAACCCCCAAACTCCTTCACTC | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | CTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACCTTCACTCTGCAGGACAG | | | | | |
| | 420 | 430 | 440 | 450 | 460 | 470 |
| ss.R02548 | 190 | 200 | 210 | 220 | | |
| | AGACGGCGGGTGCTCCTGTAGAGCGGCAAGG----- | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | AG-GGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCACCTATCA | | | | | |
| | 480 | 490 | 500 | 510 | 520 | 530 |
| ss.R02548 | ----- | | | | | |
| ss.DNA57836 | CCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGC | | | | | |
| | 540 | 550 | 560 | 570 | 580 | 590 |
| ss.R02548 | 230 | 240 | 250 | 260 | 270 | |
| | -----ATGGGGCTCTCCAGGGGACCCTGCCAGTCTCCATCTTCTGGTCACCACCT | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | AGCCTGCCAACTTCTCCTTCTGCCGAGCCAGACA-TCGGACTGGTTCTGGTGCCAGGCT | | | | | |
| | 600 | 610 | 620 | 630 | 640 | 650 |
| ss.R02548 | 280 | 290 | 300 | 310 | 320 | 330 |
| | GGGCCCA---CGTGGGTCCAGCCAGCATCCTGGGAGGTGATGGCCGAGTGGGAGGGCAA | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | GCAAACAACGCCAATGTCCAGCACAGCGCCCTCACA-GTGGTGCCCCAGGTGGTGACCA | | | | | |
| | 660 | 670 | 680 | 690 | 700 | 710 |
| ss.R02548 | 340 | 350 | 360 | 370 | | |
| | GGXTGCCAACCAGCACGA-TGGTGGGGTGCCCTGGGGAAACCC-TCCTGG | | | | | |
| | * * * * * | | | | | |
| ss.DNA57836 | G-----AAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCT | | | | | |
| | 720 | 730 | 740 | 750 | 760 | |
| ss.DNA57836 | CTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGG | | | | | |
| | 770 | 780 | 790 | 800 | 810 | 820 |
| ss.DNA57836 | GGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGACCAAGCAC | | | | | |
| | 830 | 840 | 850 | 860 | 870 | 880 |
| ss.DNA57836 | GGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAAAATGAG | | | | | |
| | 890 | 900 | 910 | 920 | 930 | 940 |
| ss.DNA57836 | TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA | | | | | |
| | 950 | 960 | 970 | 980 | 990 | |

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☒ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.